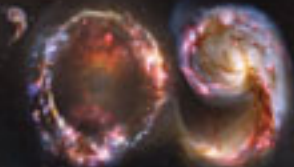


# nature



## A BIG YEAR FOR ASTRONOMY

The telescope turns 400  
Campaign for dark skies  
Hubble's greatest hits

### HIV/AIDS

A case for T-cell-based vaccines

### LEUKAEMIA STEM CELLS

Self-renewal preserved by p21 protein

### NATURE.COM

Find all our journals

# Starry messages

The first scientific observations with telescopes displaced Earth from the centre of the Universe. Modern technology continues to humble us but should not distance us from the cosmos itself.

Four hundred years ago, the Universe changed. Or, at least, our perception of it did, thanks to Galileo Galilei's scrutiny of the night sky with a telescope. Within a couple of years, his observations of the Moon, phases of Venus and satellites of Jupiter shattered the old Ptolemaic model of our Solar System. To the church's dismay, Earth assumed its rightful place as one of several planets orbiting the Sun (see page 28).

Marking Galileo's anniversary, the International Year of Astronomy seeks to remind us of the humbling nature of gazing at the heavens. Through programmes offering cheap telescopes for the masses, teaching materials and global heritage projects, astronomers are inviting us all to look up and ponder our place in the Universe.

That this prospect is novel speaks to the distance that has come between people and their environment. Urban light pollution means that one-fifth of the world's population can no longer see the Milky Way with the naked eye. Many city kids, even if they did peer through the orange smog above their heads, would probably see only a handful of stars. Astronomers are calling for the dark places on Earth to be preserved as national parks, so that we do not lose completely our window on the Universe (see page 27).

Astronomers, in communicating their work, have a natural ally in beautiful images. The Hubble Space Telescope has a place in the public's heart for this reason — although it has accomplished outstanding science too (see page 41). But professional astronomers themselves are not immune to distancing. The technological advances that have given so much insight into astrophysics have made astronomy one of the most computer-intensive of the natural sciences. Surveys churn out terabytes of data to be crunched by armies of postdocs and students



who might never have visited a mountain-top observatory. Could data mining soon replace serendipitous observational discovery? Astronomers will anyway need new visualization tools, as demonstrated by a three-dimensional graphic in this issue (see page 63), to hunt for the unexpected.

Since Galileo, advancing telescope technology has made our human perspective seem ever more myopic. We now know that the Milky Way is but one of many galaxies scattered throughout the Universe. Space is littered with fearsome objects, from giant black holes to energetic  $\gamma$ -ray bursts. Planets are plentiful around other stars. Even the normal matter that makes up our bodies is only a minority constituent of the Universe — most matter is exotic and dark.

The next generation of telescopes (see page 18) will push astronomy into a new realm. Within two decades the whole night sky will be scanned continuously and recorded digitally across many wavebands. Trips to observatories will become obsolete, except for detailed follow-up work. Astronomical data will be generated on an industrial scale and mined from a laptop. Yet the new vistas revealed will still surprise.

Every now and then, someone predicts the 'end of science'. It was believed in the 1890s that there were just a few problems left to solve in physics: yet Albert Einstein solved three of them in 1905, and founded three new branches of physics. In the next 40 years, never mind the next 400, astronomy will change. It is not the end, but the beginning of a new phase. The Universe will continue to humble us, if we take the time to look. ■

**"One-fifth of the world's population can no longer see the Milky Way with the naked eye."**

## Experts still needed

There are good reasons to be suspicious of metric-based research assessment.

Many countries are keen to measure their universities' research performance with minimal burdens on the participants. Not least of these is the United Kingdom, which last month announced the results of its sixth and final Research Assessment Exercise (RAE, see page 13).

The RAE relied heavily on expert peer review of research publications, and attention in Britain and beyond is now focused on what form the replacement system will take. The proposed successor, the Research Excellence Framework (REF), is opaque. Little is known about how it will work other than a central principle: it will assess research quality using metrics, including publication citations. It may also take into account the number of postgraduates completing their

studies and the amount of research income won by universities. There will be a smattering of 'light-touch expert review', although the exact form that this will take is not yet clear — it might simply be used to interpret the metrics results.

But taken alone, publication citations have repeatedly been shown to be a poor measure of research quality. An example from this journal illustrates the point. Our third most highly cited paper in 2007, with 272 citations at the time of inspection, was of a pilot study in screening for functional elements of the human genome. The importance lay primarily in the technique.

In contrast, a paper from the same year revealing key biological insights into the workings of a proton pump, which moves protons across cell membranes, had received 10 citations. There are plenty more examples of such large disparities between papers that may be important for a variety of reasons: technological breakthroughs of immediate use to many, more rarefied achievements of textbook status, critical insights of relevance to small

**"Metrics are not well established for the applications of science."**

# Starry messages

The first scientific observations with telescopes displaced Earth from the centre of the Universe. Modern technology continues to humble us but should not distance us from the cosmos itself.

Four hundred years ago, the Universe changed. Or, at least, our perception of it did, thanks to Galileo Galilei's scrutiny of the night sky with a telescope. Within a couple of years, his observations of the Moon, phases of Venus and satellites of Jupiter shattered the old Ptolemaic model of our Solar System. To the church's dismay, Earth assumed its rightful place as one of several planets orbiting the Sun (see page 28).

Marking Galileo's anniversary, the International Year of Astronomy seeks to remind us of the humbling nature of gazing at the heavens. Through programmes offering cheap telescopes for the masses, teaching materials and global heritage projects, astronomers are inviting us all to look up and ponder our place in the Universe.

That this prospect is novel speaks to the distance that has come between people and their environment. Urban light pollution means that one-fifth of the world's population can no longer see the Milky Way with the naked eye. Many city kids, even if they did peer through the orange smog above their heads, would probably see only a handful of stars. Astronomers are calling for the dark places on Earth to be preserved as national parks, so that we do not lose completely our window on the Universe (see page 27).

Astronomers, in communicating their work, have a natural ally in beautiful images. The Hubble Space Telescope has a place in the public's heart for this reason — although it has accomplished outstanding science too (see page 41). But professional astronomers themselves are not immune to distancing. The technological advances that have given so much insight into astrophysics have made astronomy one of the most computer-intensive of the natural sciences. Surveys churn out terabytes of data to be crunched by armies of postdocs and students



who might never have visited a mountain-top observatory. Could data mining soon replace serendipitous observational discovery? Astronomers will anyway need new visualization tools, as demonstrated by a three-dimensional graphic in this issue (see page 63), to hunt for the unexpected.

Since Galileo, advancing telescope technology has made our human perspective seem ever more myopic. We now know that the Milky Way is but one of many galaxies scattered throughout the Universe. Space is littered with fearsome objects, from giant black holes to energetic  $\gamma$ -ray bursts. Planets are plentiful around other stars. Even the normal matter that makes up our bodies is only a minority constituent of the Universe — most matter is exotic and dark.

The next generation of telescopes (see page 18) will push astronomy into a new realm. Within two decades the whole night sky will be scanned continuously and recorded digitally across many wavebands. Trips to observatories will become obsolete, except for detailed follow-up work. Astronomical data will be generated on an industrial scale and mined from a laptop. Yet the new vistas revealed will still surprise.

Every now and then, someone predicts the 'end of science'. It was believed in the 1890s that there were just a few problems left to solve in physics: yet Albert Einstein solved three of them in 1905, and founded three new branches of physics. In the next 40 years, never mind the next 400, astronomy will change. It is not the end, but the beginning of a new phase. The Universe will continue to humble us, if we take the time to look. ■

**"One-fifth of the world's population can no longer see the Milky Way with the naked eye."**

## Experts still needed

There are good reasons to be suspicious of metric-based research assessment.

Many countries are keen to measure their universities' research performance with minimal burdens on the participants. Not least of these is the United Kingdom, which last month announced the results of its sixth and final Research Assessment Exercise (RAE, see page 13).

The RAE relied heavily on expert peer review of research publications, and attention in Britain and beyond is now focused on what form the replacement system will take. The proposed successor, the Research Excellence Framework (REF), is opaque. Little is known about how it will work other than a central principle: it will assess research quality using metrics, including publication citations. It may also take into account the number of postgraduates completing their

studies and the amount of research income won by universities. There will be a smattering of 'light-touch expert review', although the exact form that this will take is not yet clear — it might simply be used to interpret the metrics results.

But taken alone, publication citations have repeatedly been shown to be a poor measure of research quality. An example from this journal illustrates the point. Our third most highly cited paper in 2007, with 272 citations at the time of inspection, was of a pilot study in screening for functional elements of the human genome. The importance lay primarily in the technique.

In contrast, a paper from the same year revealing key biological insights into the workings of a proton pump, which moves protons across cell membranes, had received 10 citations. There are plenty more examples of such large disparities between papers that may be important for a variety of reasons: technological breakthroughs of immediate use to many, more rarefied achievements of textbook status, critical insights of relevance to small

**"Metrics are not well established for the applications of science."**

or large communities, 'slow burners' whose impact grows gradually or suddenly after a delay, and so on.

Such isolated statistics serve to illustrate a point that has been more systematically documented in the bibliometrics literature. Take, for example, an analysis of the correlation between judgements of scientific value using metrics, including citations, and those using peer review, in condensed-matter physics (E. J. Rini *et al. Res. Policy* 27, 95–107; 1998). The study found disagreements in judgement between the two methods of evaluation in 25% of the 5,000 papers examined. In roughly half of these cases, the experts found a paper to be of interest when the metrics did not, and in the other half, the opposite was the case. The reasons for the differences are not fully understood.

It is also important to note that the use of metrics as an evaluation method does not have widespread support within the scientific community. Some members of the expert panels judging work in this

year's RAE warn of the dangers to the quality of research assessment under a metrics-based model. These fears were expressed even by experts in subject areas thought to be most appropriate for metrics-based assessment, such as biology and chemistry. Metrics are not well established for the applications of science, or for disciplines less dependent on journal publication.

Britain is not alone in encountering problems in developing robust metric indicators of research quality. The Australian government is also dealing with a backlash from some universities and leading researchers against its current attempts to do the same thing.

The signs are that, after several false starts and delays, the final proposals for the REF, due in autumn 2009, are unlikely to be the radical departure from the RAE that the government first envisaged in 2006. Expert review is far from a problem-free method of assessment, but policy-makers have no option but to recognize its indispensable and central role. ■

## A public service

The Christmas bird count is a model to be emulated in distributed, volunteer science.

**T**he 5th of January marks the completion of the 109th Christmas bird count, a yearly rite in which groups of North American bird-lovers pick a day around the winter solstice, fan out in teams to their designated areas, and count every bird that they see.

Held every year since 1900, when the National Audubon Society proposed it as an alternative to the then-popular Yuletide activity of competitively shooting birds, the count is the longest-running volunteer science project in the world. Its data have informed reams of peer-reviewed work, such as an ongoing effort by Audubon researchers to predict how birds will adjust their ranges in response to climate change.

The count has served as a model for any number of volunteer science efforts. Such projects now flourish — not least because the Internet makes it so easy for scientists to find, recruit and coordinate the volunteers. Out in the field, examples range from Project BudBurst, in which participants report on the timing of climate-influenced botanical events such as flowering and leafing, to the Great World Wide Star Count, in which astronomy buffs check the number of stars visible in certain bright constellations as a way of monitoring light pollution.

Indoors, meanwhile, network-based projects include Folding@home, in which millions of users allow their idle home computers to be used to simulate protein folding, and Galaxy Zoo, in which participants use their prowess at pattern recognition to classify the millions of galaxies captured in telescopic images — something that still flummoxes computers.

The lesson of this list is that the world is full of enthusiastic people — and that the opportunities for researchers to tap into this enthusiasm are limited only by their own imaginations. Volunteer science is a win-win situation for all concerned. Scientists get to take on projects that would not be feasible for even the largest research group, while helping to increase the public's understanding of, and support for,

science. And the volunteers get to have fun, while experiencing the satisfaction of defending the environment, fighting disease or expanding human knowledge.

So researchers should think creatively about whether the data they need, or the crunching or sorting they must do, can be outsourced to members of the public. And while they are at it, perhaps they should also consider joining one or more citizen science projects themselves. Participation in such efforts can reconnect scientists consumed with grant-writing and project management with the 'doing' of science. In the Christmas bird count, the most skilled bird spotters and identifiers are inevitably the non-scientists; professional ornithologists spend too much time doing paperwork. And, of course, volunteering for science feels good, especially when you see a black oystercatcher, say, or two merging galaxies — something fun, beautiful and rare. ■

### ANNOUNCEMENT

## Evolutionary gems

About a year ago, an Editorial in these pages urged scientists and their institutions to 'spread the word' and highlight reasons why scientists can treat evolution by natural selection as, in effect, an established fact (see *Nature* 451, 108; 2008).

This week we are following our own prescription. Readers will find at [www.nature.com/evolutiongems](http://www.nature.com/evolutiongems) a freely accessible resource for biologists and others who wish to explain to students, friends or loved ones just what is the evidence for evolution by natural selection. Entitled '15 evolutionary gems', the document summarizes 15 lines of evidence from papers published in *Nature* over the past 10 years. The evidence is drawn from the fossil record, from studies of natural and artificial habitats, and from research on molecular biological processes.

In a year in which Darwin is being celebrated amid uncertainty and hostility about his ideas among citizens, being aware of the cumulatively incontrovertible evidence for those ideas is all the more important. We trust that this document will help.



or large communities, 'slow burners' whose impact grows gradually or suddenly after a delay, and so on.

Such isolated statistics serve to illustrate a point that has been more systematically documented in the bibliometrics literature. Take, for example, an analysis of the correlation between judgements of scientific value using metrics, including citations, and those using peer review, in condensed-matter physics (E. J. Rini *et al. Res. Policy* 27, 95–107; 1998). The study found disagreements in judgement between the two methods of evaluation in 25% of the 5,000 papers examined. In roughly half of these cases, the experts found a paper to be of interest when the metrics did not, and in the other half, the opposite was the case. The reasons for the differences are not fully understood.

It is also important to note that the use of metrics as an evaluation method does not have widespread support within the scientific community. Some members of the expert panels judging work in this

year's RAE warn of the dangers to the quality of research assessment under a metrics-based model. These fears were expressed even by experts in subject areas thought to be most appropriate for metrics-based assessment, such as biology and chemistry. Metrics are not well established for the applications of science, or for disciplines less dependent on journal publication.

Britain is not alone in encountering problems in developing robust metric indicators of research quality. The Australian government is also dealing with a backlash from some universities and leading researchers against its current attempts to do the same thing.

The signs are that, after several false starts and delays, the final proposals for the REF, due in autumn 2009, are unlikely to be the radical departure from the RAE that the government first envisaged in 2006. Expert review is far from a problem-free method of assessment, but policy-makers have no option but to recognize its indispensable and central role. ■

## A public service

The Christmas bird count is a model to be emulated in distributed, volunteer science.

**T**he 5th of January marks the completion of the 109th Christmas bird count, a yearly rite in which groups of North American bird-lovers pick a day around the winter solstice, fan out in teams to their designated areas, and count every bird that they see.

Held every year since 1900, when the National Audubon Society proposed it as an alternative to the then-popular Yuletide activity of competitively shooting birds, the count is the longest-running volunteer science project in the world. Its data have informed reams of peer-reviewed work, such as an ongoing effort by Audubon researchers to predict how birds will adjust their ranges in response to climate change.

The count has served as a model for any number of volunteer science efforts. Such projects now flourish — not least because the Internet makes it so easy for scientists to find, recruit and coordinate the volunteers. Out in the field, examples range from Project BudBurst, in which participants report on the timing of climate-influenced botanical events such as flowering and leafing, to the Great World Wide Star Count, in which astronomy buffs check the number of stars visible in certain bright constellations as a way of monitoring light pollution.

Indoors, meanwhile, network-based projects include Folding@home, in which millions of users allow their idle home computers to be used to simulate protein folding, and Galaxy Zoo, in which participants use their prowess at pattern recognition to classify the millions of galaxies captured in telescopic images — something that still flummoxes computers.

The lesson of this list is that the world is full of enthusiastic people — and that the opportunities for researchers to tap into this enthusiasm are limited only by their own imaginations. Volunteer science is a win-win situation for all concerned. Scientists get to take on projects that would not be feasible for even the largest research group, while helping to increase the public's understanding of, and support for,

science. And the volunteers get to have fun, while experiencing the satisfaction of defending the environment, fighting disease or expanding human knowledge.

So researchers should think creatively about whether the data they need, or the crunching or sorting they must do, can be outsourced to members of the public. And while they are at it, perhaps they should also consider joining one or more citizen science projects themselves. Participation in such efforts can reconnect scientists consumed with grant-writing and project management with the 'doing' of science. In the Christmas bird count, the most skilled bird spotters and identifiers are inevitably the non-scientists; professional ornithologists spend too much time doing paperwork. And, of course, volunteering for science feels good, especially when you see a black oystercatcher, say, or two merging galaxies — something fun, beautiful and rare. ■

### ANNOUNCEMENT

## Evolutionary gems

About a year ago, an Editorial in these pages urged scientists and their institutions to 'spread the word' and highlight reasons why scientists can treat evolution by natural selection as, in effect, an established fact (see *Nature* 451, 108; 2008).

This week we are following our own prescription. Readers will find at [www.nature.com/evolutiongems](http://www.nature.com/evolutiongems) a freely accessible resource for biologists and others who wish to explain to students, friends or loved ones just what is the evidence for evolution by natural selection. Entitled '15 evolutionary gems', the document summarizes 15 lines of evidence from papers published in *Nature* over the past 10 years. The evidence is drawn from the fossil record, from studies of natural and artificial habitats, and from research on molecular biological processes.

In a year in which Darwin is being celebrated amid uncertainty and hostility about his ideas among citizens, being aware of the cumulatively incontrovertible evidence for those ideas is all the more important. We trust that this document will help.

or large communities, 'slow burners' whose impact grows gradually or suddenly after a delay, and so on.

Such isolated statistics serve to illustrate a point that has been more systematically documented in the bibliometrics literature. Take, for example, an analysis of the correlation between judgements of scientific value using metrics, including citations, and those using peer review, in condensed-matter physics (E. J. Rini *et al. Res. Policy* 27, 95–107; 1998). The study found disagreements in judgement between the two methods of evaluation in 25% of the 5,000 papers examined. In roughly half of these cases, the experts found a paper to be of interest when the metrics did not, and in the other half, the opposite was the case. The reasons for the differences are not fully understood.

It is also important to note that the use of metrics as an evaluation method does not have widespread support within the scientific community. Some members of the expert panels judging work in this

year's RAE warn of the dangers to the quality of research assessment under a metrics-based model. These fears were expressed even by experts in subject areas thought to be most appropriate for metrics-based assessment, such as biology and chemistry. Metrics are not well established for the applications of science, or for disciplines less dependent on journal publication.

Britain is not alone in encountering problems in developing robust metric indicators of research quality. The Australian government is also dealing with a backlash from some universities and leading researchers against its current attempts to do the same thing.

The signs are that, after several false starts and delays, the final proposals for the REF, due in autumn 2009, are unlikely to be the radical departure from the RAE that the government first envisaged in 2006. Expert review is far from a problem-free method of assessment, but policy-makers have no option but to recognize its indispensable and central role. ■

## A public service

The Christmas bird count is a model to be emulated in distributed, volunteer science.

**T**he 5th of January marks the completion of the 109th Christmas bird count, a yearly rite in which groups of North American bird-lovers pick a day around the winter solstice, fan out in teams to their designated areas, and count every bird that they see.

Held every year since 1900, when the National Audubon Society proposed it as an alternative to the then-popular Yuletide activity of competitively shooting birds, the count is the longest-running volunteer science project in the world. Its data have informed reams of peer-reviewed work, such as an ongoing effort by Audubon researchers to predict how birds will adjust their ranges in response to climate change.

The count has served as a model for any number of volunteer science efforts. Such projects now flourish — not least because the Internet makes it so easy for scientists to find, recruit and coordinate the volunteers. Out in the field, examples range from Project BudBurst, in which participants report on the timing of climate-influenced botanical events such as flowering and leafing, to the Great World Wide Star Count, in which astronomy buffs check the number of stars visible in certain bright constellations as a way of monitoring light pollution.

Indoors, meanwhile, network-based projects include Folding@home, in which millions of users allow their idle home computers to be used to simulate protein folding, and Galaxy Zoo, in which participants use their prowess at pattern recognition to classify the millions of galaxies captured in telescopic images — something that still flummoxes computers.

The lesson of this list is that the world is full of enthusiastic people — and that the opportunities for researchers to tap into this enthusiasm are limited only by their own imaginations. Volunteer science is a win-win situation for all concerned. Scientists get to take on projects that would not be feasible for even the largest research group, while helping to increase the public's understanding of, and support for,

science. And the volunteers get to have fun, while experiencing the satisfaction of defending the environment, fighting disease or expanding human knowledge.

So researchers should think creatively about whether the data they need, or the crunching or sorting they must do, can be outsourced to members of the public. And while they are at it, perhaps they should also consider joining one or more citizen science projects themselves. Participation in such efforts can reconnect scientists consumed with grant-writing and project management with the 'doing' of science. In the Christmas bird count, the most skilled bird spotters and identifiers are inevitably the non-scientists; professional ornithologists spend too much time doing paperwork. And, of course, volunteering for science feels good, especially when you see a black oystercatcher, say, or two merging galaxies — something fun, beautiful and rare. ■

### ANNOUNCEMENT

## Evolutionary gems

About a year ago, an Editorial in these pages urged scientists and their institutions to 'spread the word' and highlight reasons why scientists can treat evolution by natural selection as, in effect, an established fact (see *Nature* 451, 108; 2008).

This week we are following our own prescription. Readers will find at [www.nature.com/evolutiongems](http://www.nature.com/evolutiongems) a freely accessible resource for biologists and others who wish to explain to students, friends or loved ones just what is the evidence for evolution by natural selection. Entitled '15 evolutionary gems', the document summarizes 15 lines of evidence from papers published in *Nature* over the past 10 years. The evidence is drawn from the fossil record, from studies of natural and artificial habitats, and from research on molecular biological processes.

In a year in which Darwin is being celebrated amid uncertainty and hostility about his ideas among citizens, being aware of the cumulatively incontrovertible evidence for those ideas is all the more important. We trust that this document will help.

## NEWS

# Obama's picks underline climate focus

Strong roles for biologists as the president-elect chooses his science and technology team.

John Holdren, a leading voice on climate change at Harvard University, will serve as science adviser to US president-elect Barack Obama. And Jane Lubchenco, a strong advocate of marine conservation at Oregon State University in Corvallis, will head the National Oceanic and Atmospheric Administration (NOAA).

Holdren, as is normal for US science advisers, has a background in physics. At the time of his appointment, two eminent biologists were named to co-chair the President's Council of Advisors on Science and Technology (PCAST) with him. One will be Harold Varmus, the former director of the National Institutes of Health, who led Obama's science advisory team during the campaign. The other is Eric Lander, founding director of the Broad Institute in Cambridge, Massachusetts, who helped lead the push to sequence the human genome.

Together, these appointments underscore how Obama is choosing experienced academics for positions in government once he becomes president on 20 January.

If confirmed by the Senate, Holdren will replace John Marburger as head of the Office of Science and Technology Policy (OSTP), which helps set research agendas and budgets across multiple federal agencies. Holdren will also hold the title of Assistant to the President for Science and Technology; this is seen by some as a clear step up from the title Marburger enjoys, Science Advisor to the President (see *Nature* 455, 453; 2008).

Holdren's appointment quickly won praise from other academics. "He's competent and at the forefront of so many fields," says Ralph Cicerone, an atmospheric scientist and president of the National Academy of Sciences in Washington DC. "He's encyclopaedic, he's quick and he's deep." Representative Vernon Ehlers (Republican, Michigan), a physicist, adds: "He's an excellent choice, a good scientist, and I think he will serve the president and the country well."

Holdren also directs the Woods Hole Research Center in Falmouth, Massachusetts. Now an environmental-policy specialist, in his early career he worked as an engineer at Lockheed Missiles and Space Company in Sunnyvale, California, and as a fusion scientist at the Lawrence Livermore National Laboratory, also in California. He worked on nuclear weapons and non-proliferation issues while chairman of the executive committee for the Pugwash Conferences on Science and World Affairs — a familiarity with this issue that is in keeping with past science advisers. "He is comfortably in the same sort of mould that we've had for many decades," says John Gibbons, a science adviser to President Bill Clinton, "except for the unusual extent in the depth of his involvement in the process of science in government."

Holdren is no stranger to Washington, having served on the PCAST to Clinton. In that capacity, he chaired a number of major reports

meant to guide energy policy, including one that promoted a move away from coal and towards nuclear energy and renewables, says Charles Vest, president of the National Academy of Engineering. "All of the things that you expect in 2008 — but that you might not have seen on the agenda about a decade ago," says Vest.

For another PCAST report, Holdren delivered a solo presentation to Clinton on non-proliferation strategy in the early era of post-Soviet Russia, says Gibbons. "It was very coherent, expert — it answered the question authoritatively," he says. "That ultimately led to a US position on plutonium disposition a couple of weeks later when the president went to Moscow." Holdren's understanding of how the White House works will serve him well, says Neal Lane, another former Clinton science adviser.

## Rare skill

In recent years, Holdren has been a tireless advocate for improving US energy policy, travelling the world and lecturing on science's role in sustainability. He has said that reading *The Challenge of Man's Future* by Harrison Brown — which looked at the science of sustainability — in high school set him on a path to working on the intersection between science, technology and society. He has even appeared on David Letterman's late-night show to discuss the science of global warming.

Daniel Schrag, a climate scientist and Harvard colleague, says that Holdren's wide-ranging interests belie his depth of knowledge in a number of areas. "John has a remarkable ability to survey vast areas of scientific information and distil them down to their essence — that's a very rare skill," says Schrag. "He knows what he doesn't know, and he knows who to ask." Schrag also says that Holdren is skilled at getting people from different backgrounds together; as co-chairman of the National Commission on Energy Policy, Holdren led a bipartisan group including business and industry leaders to produce a document in 2004 on how to "end the energy stalemate". Holdren understands that fixing climate change comes with big costs, says Schrag. His familiarity with Lawrence Summers, the former Harvard president who will head the National Economic Council, might help him to that end.

One unresolved question is how all the climate and energy specialists in the Obama

**"He knows what he doesn't know, and he knows who to ask."**



Looking to the future: John Holdren, has long pushed for changes to US energy policy.

K. SRAKOCIC/AP PHOTO





**POLITICAL MELTDOWN**  
Canadian scientists face  
uncertain future.  
[www.nature.com/news](http://www.nature.com/news)

J. LUBCHENCO



**Marine star:** Jane Lubchenco will run the National Oceanographic and Atmospheric Administration.

administration might work together. Nobel-prizewinning physicist Steven Chu will head the Department of Energy; former Environmental Protection Agency chief Carol Browner is in a new White House position overseeing climate and energy policy. "It'll all have to be coordinated very carefully," says Lane.

Lubchenco, like Holdren, is a past president of the American Association for the Advancement of Science. A marine ecologist, she has spoken out against overfishing and done research in the hypoxic, or dead, zones that can be caused in some areas by fertilizer run-off. As head of NOAA she will have jurisdiction over a wide range of issues including the National Marine Fisheries Service and the National Weather Service.

Dealing with fisheries "will be one of her major challenges," says John Byrne, who headed NOAA during the administration of Ronald Reagan and is a former president of Oregon State University. Industry groups are wary of her pro-conservation stances, but Byrne thinks that Lubchenco is up to the challenge of running the agency. "It needs someone to respond to issues such as climate change, pressure on coastal zones, ocean pollution and over-fishing in a firm way. She will do that," he says.

Another appointment in Obama's environmental team is Ken Salazar, a Democratic senator from Colorado, to head the Department of the Interior. Salazar is known as a middle-of-the-roader who has protested against Bush administration plans to expand oil-shale development in the American west, but who has also supported offshore drilling. The agency oversees the US Fish and Wildlife Service, which has been buffeted in recent years over its handling of species listings under the Endangered Species Act. ■

**Eric Hand and Alexandra Witze, with additional reporting by Ashley Yeager.**

# Universities struggle as value of endowments falls



RECESSION  
WATCH

Some large research universities have suffered double-digit declines in their multibillion-dollar endowments since the autumn credit crunch and stock market collapse.

The endowments have also become increasingly cash-poor precisely when they might be expected to compensate for downward pressures on academic budgets. The declines are presenting universities with tough choices: fire sales of endowment assets, or budgetary trimming in the form of pay cuts, freezes on hiring and deferred construction projects.

"It's a very big problem," says John Walda, president of the National Association of College and University Business Officers (NACUBO) in Washington DC. "A good number [of endowments] are seeing about a 30% decline since last year."

Harvard University announced in December that its endowment, the world's largest and worth \$36.9 billion at mid-year, had dropped 22% from that amount by the end of November — with worse expected to come. As of mid-December, the world's second-largest endowment, that of Yale University, had fallen 25% since mid-year, when it stood at \$22.9 billion.

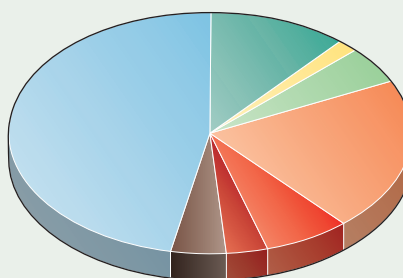
In some ways, the big universities have been a victim of their own success. As endowments experienced double-digit increases nearly every year through the 1990s and 2000s, universities have become ever more dependent on them in their overall budgets — even as endowment payouts remained roughly constant at around 5% per year. Ten years ago, Harvard's endowment paid for a third of the operating budget of the school of arts and sciences; now it covers more than half of the \$1.16-billion budget.

Endowments are suffering now partly because of the very investment methods they used to beat the broader markets year after year. Universities with endowments bigger than \$1 billion — there were 76 of them in the United States in 2007, according to NACUBO — kept much of their money in more volatile 'alternative' investments that are 'illiquid', or difficult to convert into cash quickly, such as hedge funds and venture capital. By comparison, institutions with endowments of less than \$25 million had a nearly opposite, more conservative approach, with much of their money being kept in cash and fixed-income assets (see chart).

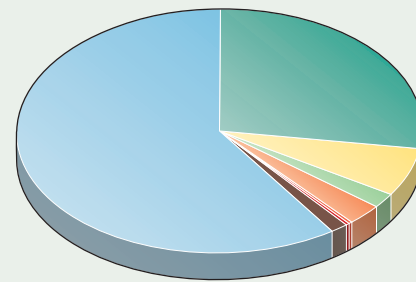
John Nelson, an analyst with Moody's in New York, says the alternative-investment approach is smart: even in the bear market, when share prices are falling, the stock

## US UNIVERSITY ENDOWMENTS: A breakdown of asset classes

Universities with investments greater than \$1 billion



Universities with investments of \$25 million or less



Equity Fixed income Cash Real estate  
Hedge funds Private equity Venture capital Other

SOURCE: NACUBO





**POLITICAL MELTDOWN**  
Canadian scientists face  
uncertain future.  
[www.nature.com/news](http://www.nature.com/news)

J. LUBCHENCO



**Marine star:** Jane Lubchenco will run the National Oceanographic and Atmospheric Administration.

administration might work together. Nobel-prizewinning physicist Steven Chu will head the Department of Energy; former Environmental Protection Agency chief Carol Browner is in a new White House position overseeing climate and energy policy. "It'll all have to be coordinated very carefully," says Lane.

Lubchenco, like Holdren, is a past president of the American Association for the Advancement of Science. A marine ecologist, she has spoken out against overfishing and done research in the hypoxic, or dead, zones that can be caused in some areas by fertilizer run-off. As head of NOAA she will have jurisdiction over a wide range of issues including the National Marine Fisheries Service and the National Weather Service.

Dealing with fisheries "will be one of her major challenges," says John Byrne, who headed NOAA during the administration of Ronald Reagan and is a former president of Oregon State University. Industry groups are wary of her pro-conservation stances, but Byrne thinks that Lubchenco is up to the challenge of running the agency. "It needs someone to respond to issues such as climate change, pressure on coastal zones, ocean pollution and over-fishing in a firm way. She will do that," he says.

Another appointment in Obama's environmental team is Ken Salazar, a Democratic senator from Colorado, to head the Department of the Interior. Salazar is known as a middle-of-the-roader who has protested against Bush administration plans to expand oil-shale development in the American west, but who has also supported offshore drilling. The agency oversees the US Fish and Wildlife Service, which has been buffeted in recent years over its handling of species listings under the Endangered Species Act. ■

**Eric Hand and Alexandra Witze, with additional reporting by Ashley Yeager.**

# Universities struggle as value of endowments falls



RECESSION  
WATCH

Some large research universities have suffered double-digit declines in their multibillion-dollar endowments since the autumn credit crunch and stock market collapse.

The endowments have also become increasingly cash-poor precisely when they might be expected to compensate for downward pressures on academic budgets. The declines are presenting universities with tough choices: fire sales of endowment assets, or budgetary trimming in the form of pay cuts, freezes on hiring and deferred construction projects.

"It's a very big problem," says John Walda, president of the National Association of College and University Business Officers (NACUBO) in Washington DC. "A good number [of endowments] are seeing about a 30% decline since last year."

Harvard University announced in December that its endowment, the world's largest and worth \$36.9 billion at mid-year, had dropped 22% from that amount by the end of November — with worse expected to come. As of mid-December, the world's second-largest endowment, that of Yale University, had fallen 25% since mid-year, when it stood at \$22.9 billion.

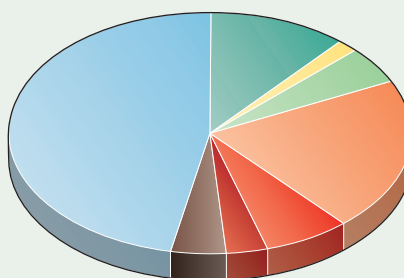
In some ways, the big universities have been a victim of their own success. As endowments experienced double-digit increases nearly every year through the 1990s and 2000s, universities have become ever more dependent on them in their overall budgets — even as endowment payouts remained roughly constant at around 5% per year. Ten years ago, Harvard's endowment paid for a third of the operating budget of the school of arts and sciences; now it covers more than half of the \$1.16-billion budget.

Endowments are suffering now partly because of the very investment methods they used to beat the broader markets year after year. Universities with endowments bigger than \$1 billion — there were 76 of them in the United States in 2007, according to NACUBO — kept much of their money in more volatile 'alternative' investments that are 'illiquid', or difficult to convert into cash quickly, such as hedge funds and venture capital. By comparison, institutions with endowments of less than \$25 million had a nearly opposite, more conservative approach, with much of their money being kept in cash and fixed-income assets (see chart).

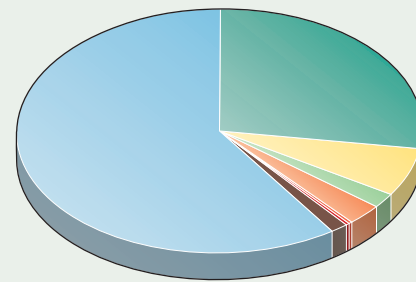
John Nelson, an analyst with Moody's in New York, says the alternative-investment approach is smart: even in the bear market, when share prices are falling, the stock

## US UNIVERSITY ENDOWMENTS: A breakdown of asset classes

Universities with investments greater than \$1 billion



Universities with investments of \$25 million or less



Equity Fixed income Cash Real estate  
Hedge funds Private equity Venture capital Other

SOURCE: NACUBO



Walking into the unknown: Harvard University's endowment has lost 22% of its value since June, and may lose more.

► portfolios of big universities have fared better than the Standard and Poor's 500 share index, which dropped 31% from 30 June to mid-December.

But alternative investments, in addition to being illiquid, can also suck up cash from elsewhere. Hedge-fund managers, for example, require periodic new investments known as 'cash calls'. These represent a problem at a time when universities need cash for salaries and construction projects and credit is hard to come by. "The tough choice that endowment managers face now is, 'when are we going to be forced to sell some of these assets in the down market?'" says Nelson. Some managers, such as those at Harvard, are already trying to sell assets off for far less than they had been worth, according to reports in *The New York Times* and *The Wall Street Journal*.

Even knowing what the alternative assets are worth is a challenge. Yeshiva University in New York City revealed that \$110 million in investments — 8% of the university's endowment — had evaporated in the investment scheme created by Bernard Madoff, who was arrested on 11 December and charged with fraud that may have cheated investors of \$50 billion (see 'Medical charity folds after investment losses', page 16). The current climate, says Walda, "will lead to more scrutiny of the underlying value of investments".

The effects on departmental hallways have been varied and will probably play out

over several years, as most endowments base their payouts on a principal amount smoothed by a three-year rolling average. Harvard has put a partial hiring freeze in place. Stanford University in Palo Alto, California, will trim its budget next year by 5% by postponing new construction, controlling salaries and cutting some jobs. The chancellor of Washington University in St Louis, Missouri, has promised to take a symbolic pay cut of nearly 10% off his half-a-million-dollar salary. The Georgia Institute of Technology in Atlanta is trimming its landscaping budget; the University of Hawaii is turning off its weekend air-conditioning; and the Field Museum of Natural History in Chicago, Illinois, is cutting its budget by 15% after the value of the museum's endowment plunged by nearly US\$100 million (31%) in the past six months (see *Nature* doi:10.1038/news.2008.1333; 2008).

### Making do

At the California Institute of Technology in Pasadena, endowment managers had started putting more money into cash and cash-equivalent assets in September, says spokesman Jon Weiner, and the university expects to weather the storm without a hiring freeze.

Few, however, are as optimistic about their prospects. In a survey released on

18 December by the National Association of Independent Colleges and Universities, half of the 371 college and university presidents who responded reported freezing new hiring. In addition, 22% reported freezing salaries, and more than two-thirds were planning to raise tuition fees for the next academic year. Their manoeuvres stem partly from falling endowments, but also from concerns such as greater difficulty in raising funds and declining student enrolment numbers.

The endowment phenomenon is largely American, where there is a long tradition of private fund-raising in higher education. In Europe, the only two universities with multibillion-dollar endowments are Oxford and Cambridge in the United Kingdom. By and large, big research universities outside the United States are much more dependent on public funding,

says Thomas Estermann of the European University Association in Brussels. "Losing a higher percentage of public funding will be much worse for them," he says. That is also the case for small universities in the United States, which are worried about budget cuts at the state level because their endowments won't be bailing them out. The 1,600 private US institutions have a median endowment of

only \$16 million, according to 2007 data from the US Department of Education. Most public universities have similarly small endowments, with the exception of a

handful of monster-sized ones, such as those of the University of Texas and the University of California systems.

Nearly a year ago, in a very different climate, universities were facing scrutiny in the US Congress. Representatives wanted to know why college tuition fees kept rising even though some universities, bolstered by successful endowments, seemed to be getting richer. They proposed legislation that would mandate state universities to spend 5% of the value of their endowment assets each year — for instance to reduce tuition fees — as is the case for private foundations. But that legislation has not yet got anywhere. Many will be watching to see if endowments, the cash cows for universities for so long, can sustain them through the long economic winter.

Eric Hand





**HAVE YOUR SAY**  
Comment on any of our  
news stories, online.  
[www.nature.com/news](http://www.nature.com/news)

# Good grades, but who gets the cash?

Britain's Research Assessment Exercise finds excellence more widespread than a focus on elite institutions would suggest.

British universities are bubbling with speculation about how the annual £1.5 billion (US\$2.2 billion) in government funding for research will be distributed, following the nation's most extensive audit of research quality.

Of the 52,400 academic researchers from 159 universities that entered the Research Assessment Exercise (RAE), published on 18 December, 17% were regarded as world-leading (rated 4\*), and 37% as internationally excellent (rated 3\*). These results will largely determine the way that higher-education funding councils allocate the money in the academic years spanning 2009–14. But universities will have to wait until 4 March 2009 to find out exactly how the results translate into cash.

The British government and the funding councils have long maintained a policy concentrating funding on the best research. Departments with high ratings are usually awarded more money per volume of staff than those with lower ranks, with funds divided according to an algorithm drawn up by the funding councils of England, Scotland and Wales. This has traditionally resulted in the same 25 or so institutions winning around 80% of the available funding, a situation that is unlikely to change this time round, according to one former vice-chancellor of a research-intensive university.

But the latest RAE results show that highly rated research is spread much more widely than that core of 25. Forty-nine universities had at least some 4\* research in their submissions, and at least half of the submissions from 118 universities were rated 4\* or 3\*. This raises fears that the funding will be spread too thinly — or that some top-quality research will not be supported. “This could seriously erode Britain's position as a world leader, particularly in biomedical sciences, in which the United Kingdom is second only to the United States,” says Steve Smith, principal of the faculty of medicine at Imperial College London.

Concentrating the funding at the top will also leave little cash for the next class of research, designated as internationally recognized (2\*), forcing funding councils to choose between their commitment to reward excellence wherever it is found, and selectively targeting their research money. “I don't think they can do both, and they now have a big problem,” the former vice-chancellor says.

“The policy is to fund the best properly and then see what is left for the rest,” confirms

a source connected to previous RAEs. He anticipates that 2\* work will receive at least some funding — there would be a great backlash otherwise, he says — but “my suspicion is that the money will certainly run out before 1\* [nationally recognized work],” he says.

Work submitted to the RAE is judged by peer review, using experts from many countries to ensure robust quality comparisons with the rest of the world. However, the RAE does not directly compare UK researchers or institutions against their overseas competitors. William Schowalter, a chemical engineer from Princeton University in New Jersey and one of the RAE's international judges, says that he is “convinced” the benchmarks were set appropriately.

“We can be confident that the results are consistent with other benchmarks indicating that the United Kingdom holds second place globally to the United States in significant subject fields,” says David Eastwood, chief executive of the Higher Education Funding Council for England (HEFCE), which runs the RAE.

*Nature* analysed RAE data on the top-performing universities in four key disciplines (see graph). The University of Cambridge ranked highest in both physics and biological sciences. And in chemistry, 40% of the university's submissions achieved a 4\* grade — 10% more than its nearest rival.

Of the four subjects, physics had the broadest distribution of top-quality research, with 16 departments awarded 4\* for 20% or more of their submissions. And biological-science

departments made more RAE submissions than any other subject, with five universities — including Leeds, Manchester and Cambridge — submitting more than 100 academics for assessment.

This year's RAE is the sixth and final exercise of its kind to be run in Britain. The government decided in 2006 to replace it with the Research Excellence Framework, which will rely more heavily on metrics such as publication citations to judge the quality of research. The move has been prompted by the expense and administrative burden that peer review places on institutions: the 2008 assessment cost HEFCE £12 million to run, more than twice as much as the previous RAE in 2001.

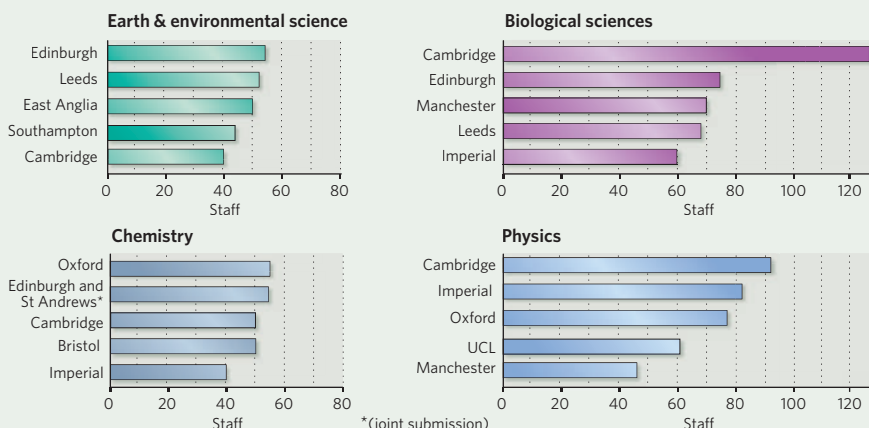
The changes will be closely watched by other nations that already use, or are in the process of implementing, similar research assessment systems, including Hong Kong, Australia and New Zealand.

A spokesman from Hong Kong's University Grants Committee, which allocates funding on the basis of the region's own RAE, told *Nature* that it is “conscious of the burden the RAE places on institutions”. The committee is already consulting academic institutions about the assessment process, and “one of the factors we shall take into account is the release of the UK RAE [2008 results] and the reaction to it, as well as further developments in the metrics model the United Kingdom is moving towards”.

**Natasha Gilbert**  
See Editorial, page 7.

## TOP OF THE CLASS

The number of staff judged ‘world leading’ (4\*) or ‘internationally excellent’ (3\*) by the 2008 Research Assessment Exercise, at the leading UK universities in four core subject areas



SOURCE: RAE

# Single fund for African science aid

Cash pool could break power struggles between competing agencies on the continent.

Plans to funnel much of the science, technology and education aid bound for Africa into a single fund are taking shape. Such a move could boost the flow of cash from donors and improve coordination of science development activities on the continent.

But critics say that the proposed central fund — which could manage billions of dollars in aid — is simply being used as a pawn in a power struggle between agencies coordinating development spending on the continent.

Since an African science and technology fund was mooted in an action plan adopted by the continent's science ministers in 2005, progress to that end has been slow. However, a pan-African science-policy meeting in Abuja, Nigeria, on 3–5 December heard that the African Development Bank is investigating whether it can host the fund, with a decision expected around March 2009.

Hakim Elwaer heads the science directorate of the African Union (AU), the confederation representing 53 African states. He told the Abuja meeting that the lack of a transparent mechanism for African funding is a key reason why many millions of dollars pledged for science by donors, such as the G8 (Group of 8) leading industrialized nations, have been slow to reach African researchers.

"The moment you have a fund structure, with an efficient, transparent and credible management system, this gives more credibility for donors to put money in — especially the big development partners from Europe, Japan, America and the World Bank," he says. He expects the majority of the contributions to the centralized fund to come from governmental and intergovernmental donors, aid agencies and African countries themselves, and emphasizes that the fund will not replace bilateral funding arrangements with individual countries.

A centralized funding system could strengthen Africa's role in defining its own development priorities, says Myles Wickstead, senior adviser with the Africa Unit of the Association of Commonwealth Universities and visiting professor of international relations at the Open University in Milton Keynes, UK.

And Elwaer hopes that the fund could also



African Union members could get development aid through a central fund.

improve coordination of science development activities on the continent, which range from financial support of individual research projects to science-literacy programmes in African parliaments. This should counter an increasing duplication of efforts as intergovernmental agencies, eager to capitalize on the willingness of donors to spend money on African science, have drawn up science strategies independently.

## Astonishing duplication

Researchers from the Network of African Science Academies are already working on identifying the overlaps between the three biggest science strategies: the 2005 action plan nurtured by the AU and its implementing partner, the New Partnership for Africa's Development (NEPAD); a second drawn up by the United Nations Economic Commission for Africa; and a third produced by the African Development Bank.

"We were absolutely astonished that these three major initiatives had been developed rather independently," says Mohammed Hassan, executive director of the Academy of Sciences for the Developing World.

But Calestous Juma, a Kenyan professor of international development at Harvard University, says that efforts to coordinate development activities are not tackling the root of the problem: big agencies are jostling to set the science agenda on the continent.

"Is duplication the reason why little is being achieved? I don't think so," he says. "Many of the organizations using this language are

simply seeking to control financial flows, not necessarily to get anything done."

Elwaer rejects the notion that the AU is exerting improper control over Africa's science agenda. But there are certainly tensions between the AU and NEPAD that have hampered progress on their joint strategy. In November 2007, a growing dispute between the AU science directorate and the NEPAD science office went public when officials engaged in a vocal row during a ministerial meeting in Mombasa, Kenya, over rival efforts to measure science

activity on the continent.

One reason for the tensions is historical. NEPAD was created in 2001, with the encouragement of G8 donors, whereas the AU was established in 2002 and took several years to find its feet. By the time that the AU was put in charge of policy, with NEPAD handling implementation, NEPAD had already spent years doing both. "We've never understood if the AU or NEPAD should take the lead," says Jimmy Whitworth, director of international activities at the Wellcome Trust, a UK medical-research charity.

But in a closed meeting at the Abuja conference, representatives of the two agencies met to try to bury the hatchet. "The meeting in Abuja seems to have been marked by a recognition that the AU and NEPAD have to work together," says Wickstead.

That should significantly improve the prospects of the African Development Bank fund getting off the ground, although there are many hurdles still to clear. By mid-2009, the AU hopes to have a clearer idea of how the fund would be managed and what sorts of programmes it would support. This would then be discussed and signed off by African science and education ministers.

And there is the added pressure of the economic downturn, which could cause donor contributions to dry up. "The timing is awkward," admits Sibry Tapsoba, the African Development Bank manager for science. "Because of the financial crisis, everybody could be saying that resources are lacking. We are very positive, but we are also very cautious."

**Linda Nordinling writes about African science policy for SciDev.Net.**

A. WAGUIH/REUTERS



## Medical charity folds after investment losses

The Picower Foundation, a large US charity that funds medical research, has collapsed after losing money in a fraudulent US\$50-billion (£34-million) investment scheme run by businessman Bernard Madoff.

The Florida-based foundation was set up in 1989 and had distributed a total of \$268 million in grants, including a \$50-million donation in 2002 to fund the Picower Institute for Learning and Memory at the Massachusetts Institute of Technology in Cambridge. The foundation had assets of \$1 billion invested in Madoff's scheme, and it is unclear whether any of these funds will be recovered.

Other organizations hit by the fraud include Tufts University in Medford, Massachusetts, which had put \$20 million into the scheme, and the American Technion Society, a science-education charity in New York, which had made a \$72-million investment.

## Spanish solar companies slammed over subsidy fraud

Many of the solar parks that stretch across the Spanish countryside are guilty of fraud, Spain's National Energy Commission (CNE) has found.

In the past two years, Spain's solar industry has grown by 900%. The country now has the third-largest solar capacity in the world, behind the United States and Germany. But the CNE's investigation has found that nearly 4,200 photovoltaic installations, found on 13% of Spain's solar parks, were falsely registered as producing power by 30 September 2008 in order to receive greater subsidies from power companies.

Industry analysts are concerned that the fraudulent installations have effectively

## First antenna switches on in the Atacama

The Atacama Large Millimeter/submillimeter Array (ALMA) — a radio telescope — has received the first of its 12-metre-wide radio antennas (pictured). The project is a multinational collaboration with a cost well in excess of US\$1 billion. It will eventually consist of 54 identical antennas — along with 12 seven-metre antennas — at a site 5,000 metres above sea level in Chile's Atacama Desert.

ALMA will be used to study the Universe at millimetre and submillimetre wavelengths, providing a new window on the cool star-forming regions of the Milky Way and on early galaxies. Further antennas should be delivered throughout 2009, and the entire array is expected to be completed by the end of 2012 (see page 18).



ALMA (ESO/NAOJ/NRAO)

used up the subsidies anticipated for 2009 — and on 17 December the government postponed a round of awards planned for January until 1 March.

For a longer version of this story, see <http://tinyurl.com/9nxa5h>.

## US firms power up to build advanced batteries

Fourteen US companies have formed a coalition to make advanced lithium-ion batteries for vehicles. The National Alliance for Advanced Transportation Battery Cell Manufacture is proposing to build one or more large manufacturing plants that would be shared by its members.

Members of the alliance include Altair Nanotechnologies of Reno, Nevada, 3M of Saint Paul, Minnesota, and Johnson Controls-Saft, a joint American-French battery-production enterprise. The US Department of Energy's Argonne National Laboratory in Illinois will have an advisory role. The alliance is seeking government support for the majority of the venture, estimated to cost between \$1 billion and \$2 billion over five years.

## Health organization lays plans for major biobank

Health-care provider Kaiser Permanente based in Oakland, California, is building a biobank that could rival the world's largest in size and scope. The organization has received a grant of US\$8.6 million from the philanthropic Robert Wood Johnson Foundation in Princeton, New Jersey, to collect DNA specimens from 200,000 people enrolled in Kaiser's medical-care plan.

Kaiser will link the DNA to participants'

existing electronic medical records and to a planned database of environmental information, with the aim of reaching a total of 500,000 participants by 2012 — making it much larger than any other current biobank in the United States. The data will be available to researchers around the world for studying how genes and other factors influence disease risk and health.

For a longer version of this story, see <http://tinyurl.com/9bvr5d>.

## NASA awards contracts to supply space station

Orbital Sciences of Dulles, Virginia, and Space Exploration Technologies (SpaceX) of Hawthorne, California, have secured contracts from NASA to ferry cargo to the International Space Station (ISS) from 2010.

The contracts, which could each be worth up to US\$3.1 billion, will see the private carriers supply 40–70% of the space station's US cargo each year until the end of 2016. The firms will be stepping into the breach left when NASA retires its space shuttles in 2010.

In September 2008, SpaceX launched the first privately developed liquid-fuel rocket and its cargo into low Earth orbit. The company's ISS deliveries will probably be made with Falcon 9, a larger version of the rocket, which is expected to make its first launch within months.

### Correction

The News Feature 'Phoenix: a race against time' (*Nature* **456**, 690–695, 2008) suggested that Honeybee Robotics failed to follow revised designs that would have corrected a flaw in a part of the NASA Phoenix spacecraft. A subsequent investigation by mission managers at NASA's Jet Propulsion Laboratory found that Honeybee was not at fault. See Correspondence, page 26.



Subsidies may run short for Spain's solar parks.

## Medical charity folds after investment losses

The Picower Foundation, a large US charity that funds medical research, has collapsed after losing money in a fraudulent US\$50-billion (£34-million) investment scheme run by businessman Bernard Madoff.

The Florida-based foundation was set up in 1989 and had distributed a total of \$268 million in grants, including a \$50-million donation in 2002 to fund the Picower Institute for Learning and Memory at the Massachusetts Institute of Technology in Cambridge. The foundation had assets of \$1 billion invested in Madoff's scheme, and it is unclear whether any of these funds will be recovered.

Other organizations hit by the fraud include Tufts University in Medford, Massachusetts, which had put \$20 million into the scheme, and the American Technion Society, a science-education charity in New York, which had made a \$72-million investment.

## Spanish solar companies slammed over subsidy fraud

Many of the solar parks that stretch across the Spanish countryside are guilty of fraud, Spain's National Energy Commission (CNE) has found.

In the past two years, Spain's solar industry has grown by 900%. The country now has the third-largest solar capacity in the world, behind the United States and Germany. But the CNE's investigation has found that nearly 4,200 photovoltaic installations, found on 13% of Spain's solar parks, were falsely registered as producing power by 30 September 2008 in order to receive greater subsidies from power companies.

Industry analysts are concerned that the fraudulent installations have effectively

## First antenna switches on in the Atacama

The Atacama Large Millimeter/submillimeter Array (ALMA) — a radio telescope — has received the first of its 12-metre-wide radio antennas (pictured). The project is a multinational collaboration with a cost well in excess of US\$1 billion. It will eventually consist of 54 identical antennas — along with 12 seven-metre antennas — at a site 5,000 metres above sea level in Chile's Atacama Desert.

ALMA will be used to study the Universe at millimetre and submillimetre wavelengths, providing a new window on the cool star-forming regions of the Milky Way and on early galaxies. Further antennas should be delivered throughout 2009, and the entire array is expected to be completed by the end of 2012 (see page 18).



ALMA (ESO/NAOJ/NRAO)

used up the subsidies anticipated for 2009 — and on 17 December the government postponed a round of awards planned for January until 1 March.

For a longer version of this story, see <http://tinyurl.com/9nxa5h>.

## US firms power up to build advanced batteries

Fourteen US companies have formed a coalition to make advanced lithium-ion batteries for vehicles. The National Alliance for Advanced Transportation Battery Cell Manufacture is proposing to build one or more large manufacturing plants that would be shared by its members.

Members of the alliance include Altair Nanotechnologies of Reno, Nevada, 3M of Saint Paul, Minnesota, and Johnson Controls-Saft, a joint American-French battery-production enterprise. The US Department of Energy's Argonne National Laboratory in Illinois will have an advisory role. The alliance is seeking government support for the majority of the venture, estimated to cost between \$1 billion and \$2 billion over five years.

## Health organization lays plans for major biobank

Health-care provider Kaiser Permanente based in Oakland, California, is building a biobank that could rival the world's largest in size and scope. The organization has received a grant of US\$8.6 million from the philanthropic Robert Wood Johnson Foundation in Princeton, New Jersey, to collect DNA specimens from 200,000 people enrolled in Kaiser's medical-care plan.

Kaiser will link the DNA to participants'

existing electronic medical records and to a planned database of environmental information, with the aim of reaching a total of 500,000 participants by 2012 — making it much larger than any other current biobank in the United States. The data will be available to researchers around the world for studying how genes and other factors influence disease risk and health.

For a longer version of this story, see <http://tinyurl.com/9bvr5d>.

## NASA awards contracts to supply space station

Orbital Sciences of Dulles, Virginia, and Space Exploration Technologies (SpaceX) of Hawthorne, California, have secured contracts from NASA to ferry cargo to the International Space Station (ISS) from 2010.

The contracts, which could each be worth up to US\$3.1 billion, will see the private carriers supply 40–70% of the space station's US cargo each year until the end of 2016. The firms will be stepping into the breach left when NASA retires its space shuttles in 2010.

In September 2008, SpaceX launched the first privately developed liquid-fuel rocket and its cargo into low Earth orbit. The company's ISS deliveries will probably be made with Falcon 9, a larger version of the rocket, which is expected to make its first launch within months.

### Correction

The News Feature 'Phoenix: a race against time' (*Nature* **456**, 690–695, 2008) suggested that Honeybee Robotics failed to follow revised designs that would have corrected a flaw in a part of the NASA Phoenix spacecraft. A subsequent investigation by mission managers at NASA's Jet Propulsion Laboratory found that Honeybee was not at fault. See Correspondence, page 26.



Subsidies may run short for Spain's solar parks.



## Medical charity folds after investment losses

The Picower Foundation, a large US charity that funds medical research, has collapsed after losing money in a fraudulent US\$50-billion (£34-million) investment scheme run by businessman Bernard Madoff.

The Florida-based foundation was set up in 1989 and had distributed a total of \$268 million in grants, including a \$50-million donation in 2002 to fund the Picower Institute for Learning and Memory at the Massachusetts Institute of Technology in Cambridge. The foundation had assets of \$1 billion invested in Madoff's scheme, and it is unclear whether any of these funds will be recovered.

Other organizations hit by the fraud include Tufts University in Medford, Massachusetts, which had put \$20 million into the scheme, and the American Technion Society, a science-education charity in New York, which had made a \$72-million investment.

## Spanish solar companies slammed over subsidy fraud

Many of the solar parks that stretch across the Spanish countryside are guilty of fraud, Spain's National Energy Commission (CNE) has found.

In the past two years, Spain's solar industry has grown by 900%. The country now has the third-largest solar capacity in the world, behind the United States and Germany. But the CNE's investigation has found that nearly 4,200 photovoltaic installations, found on 13% of Spain's solar parks, were falsely registered as producing power by 30 September 2008 in order to receive greater subsidies from power companies.

Industry analysts are concerned that the fraudulent installations have effectively

## First antenna switches on in the Atacama

The Atacama Large Millimeter/submillimeter Array (ALMA) — a radio telescope — has received the first of its 12-metre-wide radio antennas (pictured). The project is a multinational collaboration with a cost well in excess of US\$1 billion. It will eventually consist of 54 identical antennas — along with 12 seven-metre antennas — at a site 5,000 metres above sea level in Chile's Atacama Desert.

ALMA will be used to study the Universe at millimetre and submillimetre wavelengths, providing a new window on the cool star-forming regions of the Milky Way and on early galaxies. Further antennas should be delivered throughout 2009, and the entire array is expected to be completed by the end of 2012 (see page 18).



ALMA (ESO/NAOJ/NRAO)

used up the subsidies anticipated for 2009 — and on 17 December the government postponed a round of awards planned for January until 1 March.

For a longer version of this story, see <http://tinyurl.com/9nxa5h>.

## US firms power up to build advanced batteries

Fourteen US companies have formed a coalition to make advanced lithium-ion batteries for vehicles. The National Alliance for Advanced Transportation Battery Cell Manufacture is proposing to build one or more large manufacturing plants that would be shared by its members.

Members of the alliance include Altair Nanotechnologies of Reno, Nevada, 3M of Saint Paul, Minnesota, and Johnson Controls-Saft, a joint American-French battery-production enterprise. The US Department of Energy's Argonne National Laboratory in Illinois will have an advisory role. The alliance is seeking government support for the majority of the venture, estimated to cost between \$1 billion and \$2 billion over five years.

## Health organization lays plans for major biobank

Health-care provider Kaiser Permanente based in Oakland, California, is building a biobank that could rival the world's largest in size and scope. The organization has received a grant of US\$8.6 million from the philanthropic Robert Wood Johnson Foundation in Princeton, New Jersey, to collect DNA specimens from 200,000 people enrolled in Kaiser's medical-care plan.

Kaiser will link the DNA to participants'

existing electronic medical records and to a planned database of environmental information, with the aim of reaching a total of 500,000 participants by 2012 — making it much larger than any other current biobank in the United States. The data will be available to researchers around the world for studying how genes and other factors influence disease risk and health.

For a longer version of this story, see <http://tinyurl.com/9bvr5d>.

## NASA awards contracts to supply space station

Orbital Sciences of Dulles, Virginia, and Space Exploration Technologies (SpaceX) of Hawthorne, California, have secured contracts from NASA to ferry cargo to the International Space Station (ISS) from 2010.

The contracts, which could each be worth up to US\$3.1 billion, will see the private carriers supply 40–70% of the space station's US cargo each year until the end of 2016. The firms will be stepping into the breach left when NASA retires its space shuttles in 2010.

In September 2008, SpaceX launched the first privately developed liquid-fuel rocket and its cargo into low Earth orbit. The company's ISS deliveries will probably be made with Falcon 9, a larger version of the rocket, which is expected to make its first launch within months.

### Correction

The News Feature 'Phoenix: a race against time' (*Nature* **456**, 690–695, 2008) suggested that Honeybee Robotics failed to follow revised designs that would have corrected a flaw in a part of the NASA Phoenix spacecraft. A subsequent investigation by mission managers at NASA's Jet Propulsion Laboratory found that Honeybee was not at fault. See Correspondence, page 26.



Subsidies may run short for Spain's solar parks.

## Medical charity folds after investment losses

The Picower Foundation, a large US charity that funds medical research, has collapsed after losing money in a fraudulent US\$50-billion (£34-million) investment scheme run by businessman Bernard Madoff.

The Florida-based foundation was set up in 1989 and had distributed a total of \$268 million in grants, including a \$50-million donation in 2002 to fund the Picower Institute for Learning and Memory at the Massachusetts Institute of Technology in Cambridge. The foundation had assets of \$1 billion invested in Madoff's scheme, and it is unclear whether any of these funds will be recovered.

Other organizations hit by the fraud include Tufts University in Medford, Massachusetts, which had put \$20 million into the scheme, and the American Technion Society, a science-education charity in New York, which had made a \$72-million investment.

## Spanish solar companies slammed over subsidy fraud

Many of the solar parks that stretch across the Spanish countryside are guilty of fraud, Spain's National Energy Commission (CNE) has found.

In the past two years, Spain's solar industry has grown by 900%. The country now has the third-largest solar capacity in the world, behind the United States and Germany. But the CNE's investigation has found that nearly 4,200 photovoltaic installations, found on 13% of Spain's solar parks, were falsely registered as producing power by 30 September 2008 in order to receive greater subsidies from power companies.

Industry analysts are concerned that the fraudulent installations have effectively

## First antenna switches on in the Atacama

The Atacama Large Millimeter/submillimeter Array (ALMA) — a radio telescope — has received the first of its 12-metre-wide radio antennas (pictured). The project is a multinational collaboration with a cost well in excess of US\$1 billion. It will eventually consist of 54 identical antennas — along with 12 seven-metre antennas — at a site 5,000 metres above sea level in Chile's Atacama Desert.

ALMA will be used to study the Universe at millimetre and submillimetre wavelengths, providing a new window on the cool star-forming regions of the Milky Way and on early galaxies. Further antennas should be delivered throughout 2009, and the entire array is expected to be completed by the end of 2012 (see page 18).



ALMA (ESO/NAOJ/NRAO)

used up the subsidies anticipated for 2009 — and on 17 December the government postponed a round of awards planned for January until 1 March.

For a longer version of this story, see <http://tinyurl.com/9nxa5h>.

## US firms power up to build advanced batteries

Fourteen US companies have formed a coalition to make advanced lithium-ion batteries for vehicles. The National Alliance for Advanced Transportation Battery Cell Manufacture is proposing to build one or more large manufacturing plants that would be shared by its members.

Members of the alliance include Altair Nanotechnologies of Reno, Nevada, 3M of Saint Paul, Minnesota, and Johnson Controls-Saft, a joint American-French battery-production enterprise. The US Department of Energy's Argonne National Laboratory in Illinois will have an advisory role. The alliance is seeking government support for the majority of the venture, estimated to cost between \$1 billion and \$2 billion over five years.

## Health organization lays plans for major biobank

Health-care provider Kaiser Permanente based in Oakland, California, is building a biobank that could rival the world's largest in size and scope. The organization has received a grant of US\$8.6 million from the philanthropic Robert Wood Johnson Foundation in Princeton, New Jersey, to collect DNA specimens from 200,000 people enrolled in Kaiser's medical-care plan.

Kaiser will link the DNA to participants'

existing electronic medical records and to a planned database of environmental information, with the aim of reaching a total of 500,000 participants by 2012 — making it much larger than any other current biobank in the United States. The data will be available to researchers around the world for studying how genes and other factors influence disease risk and health.

For a longer version of this story, see <http://tinyurl.com/9bvr5d>.

## NASA awards contracts to supply space station

Orbital Sciences of Dulles, Virginia, and Space Exploration Technologies (SpaceX) of Hawthorne, California, have secured contracts from NASA to ferry cargo to the International Space Station (ISS) from 2010.

The contracts, which could each be worth up to US\$3.1 billion, will see the private carriers supply 40–70% of the space station's US cargo each year until the end of 2016. The firms will be stepping into the breach left when NASA retires its space shuttles in 2010.

In September 2008, SpaceX launched the first privately developed liquid-fuel rocket and its cargo into low Earth orbit. The company's ISS deliveries will probably be made with Falcon 9, a larger version of the rocket, which is expected to make its first launch within months.

### Correction

The News Feature 'Phoenix: a race against time' (*Nature* **456**, 690–695, 2008) suggested that Honeybee Robotics failed to follow revised designs that would have corrected a flaw in a part of the NASA Phoenix spacecraft. A subsequent investigation by mission managers at NASA's Jet Propulsion Laboratory found that Honeybee was not at fault. See Correspondence, page 26.



Subsidies may run short for Spain's solar parks.



## Medical charity folds after investment losses

The Picower Foundation, a large US charity that funds medical research, has collapsed after losing money in a fraudulent US\$50-billion (£34-million) investment scheme run by businessman Bernard Madoff.

The Florida-based foundation was set up in 1989 and had distributed a total of \$268 million in grants, including a \$50-million donation in 2002 to fund the Picower Institute for Learning and Memory at the Massachusetts Institute of Technology in Cambridge. The foundation had assets of \$1 billion invested in Madoff's scheme, and it is unclear whether any of these funds will be recovered.

Other organizations hit by the fraud include Tufts University in Medford, Massachusetts, which had put \$20 million into the scheme, and the American Technion Society, a science-education charity in New York, which had made a \$72-million investment.

## Spanish solar companies slammed over subsidy fraud

Many of the solar parks that stretch across the Spanish countryside are guilty of fraud, Spain's National Energy Commission (CNE) has found.

In the past two years, Spain's solar industry has grown by 900%. The country now has the third-largest solar capacity in the world, behind the United States and Germany. But the CNE's investigation has found that nearly 4,200 photovoltaic installations, found on 13% of Spain's solar parks, were falsely registered as producing power by 30 September 2008 in order to receive greater subsidies from power companies.

Industry analysts are concerned that the fraudulent installations have effectively

## First antenna switches on in the Atacama

The Atacama Large Millimeter/submillimeter Array (ALMA) — a radio telescope — has received the first of its 12-metre-wide radio antennas (pictured). The project is a multinational collaboration with a cost well in excess of US\$1 billion. It will eventually consist of 54 identical antennas — along with 12 seven-metre antennas — at a site 5,000 metres above sea level in Chile's Atacama Desert.

ALMA will be used to study the Universe at millimetre and submillimetre wavelengths, providing a new window on the cool star-forming regions of the Milky Way and on early galaxies. Further antennas should be delivered throughout 2009, and the entire array is expected to be completed by the end of 2012 (see page 18).



ALMA (ESO/NAOJ/NRAO)

used up the subsidies anticipated for 2009 — and on 17 December the government postponed a round of awards planned for January until 1 March.

For a longer version of this story, see <http://tinyurl.com/9nxa5h>.

## US firms power up to build advanced batteries

Fourteen US companies have formed a coalition to make advanced lithium-ion batteries for vehicles. The National Alliance for Advanced Transportation Battery Cell Manufacture is proposing to build one or more large manufacturing plants that would be shared by its members.

Members of the alliance include Altair Nanotechnologies of Reno, Nevada, 3M of Saint Paul, Minnesota, and Johnson Controls-Saft, a joint American-French battery-production enterprise. The US Department of Energy's Argonne National Laboratory in Illinois will have an advisory role. The alliance is seeking government support for the majority of the venture, estimated to cost between \$1 billion and \$2 billion over five years.

## Health organization lays plans for major biobank

Health-care provider Kaiser Permanente based in Oakland, California, is building a biobank that could rival the world's largest in size and scope. The organization has received a grant of US\$8.6 million from the philanthropic Robert Wood Johnson Foundation in Princeton, New Jersey, to collect DNA specimens from 200,000 people enrolled in Kaiser's medical-care plan.

Kaiser will link the DNA to participants'

existing electronic medical records and to a planned database of environmental information, with the aim of reaching a total of 500,000 participants by 2012 — making it much larger than any other current biobank in the United States. The data will be available to researchers around the world for studying how genes and other factors influence disease risk and health.

For a longer version of this story, see <http://tinyurl.com/9bvr5d>.

## NASA awards contracts to supply space station

Orbital Sciences of Dulles, Virginia, and Space Exploration Technologies (SpaceX) of Hawthorne, California, have secured contracts from NASA to ferry cargo to the International Space Station (ISS) from 2010.

The contracts, which could each be worth up to US\$3.1 billion, will see the private carriers supply 40–70% of the space station's US cargo each year until the end of 2016. The firms will be stepping into the breach left when NASA retires its space shuttles in 2010.

In September 2008, SpaceX launched the first privately developed liquid-fuel rocket and its cargo into low Earth orbit. The company's ISS deliveries will probably be made with Falcon 9, a larger version of the rocket, which is expected to make its first launch within months.

### Correction

The News Feature 'Phoenix: a race against time' (*Nature* **456**, 690–695, 2008) suggested that Honeybee Robotics failed to follow revised designs that would have corrected a flaw in a part of the NASA Phoenix spacecraft. A subsequent investigation by mission managers at NASA's Jet Propulsion Laboratory found that Honeybee was not at fault. See Correspondence, page 26.



Subsidies may run short for Spain's solar parks.

## Medical charity folds after investment losses

The Picower Foundation, a large US charity that funds medical research, has collapsed after losing money in a fraudulent US\$50-billion (£34-million) investment scheme run by businessman Bernard Madoff.

The Florida-based foundation was set up in 1989 and had distributed a total of \$268 million in grants, including a \$50-million donation in 2002 to fund the Picower Institute for Learning and Memory at the Massachusetts Institute of Technology in Cambridge. The foundation had assets of \$1 billion invested in Madoff's scheme, and it is unclear whether any of these funds will be recovered.

Other organizations hit by the fraud include Tufts University in Medford, Massachusetts, which had put \$20 million into the scheme, and the American Technion Society, a science-education charity in New York, which had made a \$72-million investment.

## Spanish solar companies slammed over subsidy fraud

Many of the solar parks that stretch across the Spanish countryside are guilty of fraud, Spain's National Energy Commission (CNE) has found.

In the past two years, Spain's solar industry has grown by 900%. The country now has the third-largest solar capacity in the world, behind the United States and Germany. But the CNE's investigation has found that nearly 4,200 photovoltaic installations, found on 13% of Spain's solar parks, were falsely registered as producing power by 30 September 2008 in order to receive greater subsidies from power companies.

Industry analysts are concerned that the fraudulent installations have effectively

## First antenna switches on in the Atacama

The Atacama Large Millimeter/submillimeter Array (ALMA) — a radio telescope — has received the first of its 12-metre-wide radio antennas (pictured). The project is a multinational collaboration with a cost well in excess of US\$1 billion. It will eventually consist of 54 identical antennas — along with 12 seven-metre antennas — at a site 5,000 metres above sea level in Chile's Atacama Desert.

ALMA will be used to study the Universe at millimetre and submillimetre wavelengths, providing a new window on the cool star-forming regions of the Milky Way and on early galaxies. Further antennas should be delivered throughout 2009, and the entire array is expected to be completed by the end of 2012 (see page 18).



ALMA (ESO/NAOJ/NRAO)

used up the subsidies anticipated for 2009 — and on 17 December the government postponed a round of awards planned for January until 1 March.

For a longer version of this story, see <http://tinyurl.com/9nxa5h>.

## US firms power up to build advanced batteries

Fourteen US companies have formed a coalition to make advanced lithium-ion batteries for vehicles. The National Alliance for Advanced Transportation Battery Cell Manufacture is proposing to build one or more large manufacturing plants that would be shared by its members.

Members of the alliance include Altair Nanotechnologies of Reno, Nevada, 3M of Saint Paul, Minnesota, and Johnson Controls-Saft, a joint American-French battery-production enterprise. The US Department of Energy's Argonne National Laboratory in Illinois will have an advisory role. The alliance is seeking government support for the majority of the venture, estimated to cost between \$1 billion and \$2 billion over five years.

## Health organization lays plans for major biobank

Health-care provider Kaiser Permanente based in Oakland, California, is building a biobank that could rival the world's largest in size and scope. The organization has received a grant of US\$8.6 million from the philanthropic Robert Wood Johnson Foundation in Princeton, New Jersey, to collect DNA specimens from 200,000 people enrolled in Kaiser's medical-care plan.

Kaiser will link the DNA to participants'

existing electronic medical records and to a planned database of environmental information, with the aim of reaching a total of 500,000 participants by 2012 — making it much larger than any other current biobank in the United States. The data will be available to researchers around the world for studying how genes and other factors influence disease risk and health.

For a longer version of this story, see <http://tinyurl.com/9bvr5d>.

## NASA awards contracts to supply space station

Orbital Sciences of Dulles, Virginia, and Space Exploration Technologies (SpaceX) of Hawthorne, California, have secured contracts from NASA to ferry cargo to the International Space Station (ISS) from 2010.

The contracts, which could each be worth up to US\$3.1 billion, will see the private carriers supply 40–70% of the space station's US cargo each year until the end of 2016. The firms will be stepping into the breach left when NASA retires its space shuttles in 2010.

In September 2008, SpaceX launched the first privately developed liquid-fuel rocket and its cargo into low Earth orbit. The company's ISS deliveries will probably be made with Falcon 9, a larger version of the rocket, which is expected to make its first launch within months.

### Correction

The News Feature 'Phoenix: a race against time' (*Nature* **456**, 690–695, 2008) suggested that Honeybee Robotics failed to follow revised designs that would have corrected a flaw in a part of the NASA Phoenix spacecraft. A subsequent investigation by mission managers at NASA's Jet Propulsion Laboratory found that Honeybee was not at fault. See Correspondence, page 26.



Subsidies may run short for Spain's solar parks.



## Medical charity folds after investment losses

The Picower Foundation, a large US charity that funds medical research, has collapsed after losing money in a fraudulent US\$50-billion (£34-million) investment scheme run by businessman Bernard Madoff.

The Florida-based foundation was set up in 1989 and had distributed a total of \$268 million in grants, including a \$50-million donation in 2002 to fund the Picower Institute for Learning and Memory at the Massachusetts Institute of Technology in Cambridge. The foundation had assets of \$1 billion invested in Madoff's scheme, and it is unclear whether any of these funds will be recovered.

Other organizations hit by the fraud include Tufts University in Medford, Massachusetts, which had put \$20 million into the scheme, and the American Technion Society, a science-education charity in New York, which had made a \$72-million investment.

## Spanish solar companies slammed over subsidy fraud

Many of the solar parks that stretch across the Spanish countryside are guilty of fraud, Spain's National Energy Commission (CNE) has found.

In the past two years, Spain's solar industry has grown by 900%. The country now has the third-largest solar capacity in the world, behind the United States and Germany. But the CNE's investigation has found that nearly 4,200 photovoltaic installations, found on 13% of Spain's solar parks, were falsely registered as producing power by 30 September 2008 in order to receive greater subsidies from power companies.

Industry analysts are concerned that the fraudulent installations have effectively

## First antenna switches on in the Atacama

The Atacama Large Millimeter/submillimeter Array (ALMA) — a radio telescope — has received the first of its 12-metre-wide radio antennas (pictured). The project is a multinational collaboration with a cost well in excess of US\$1 billion. It will eventually consist of 54 identical antennas — along with 12 seven-metre antennas — at a site 5,000 metres above sea level in Chile's Atacama Desert.

ALMA will be used to study the Universe at millimetre and submillimetre wavelengths, providing a new window on the cool star-forming regions of the Milky Way and on early galaxies. Further antennas should be delivered throughout 2009, and the entire array is expected to be completed by the end of 2012 (see page 18).



ALMA (ESO/NAOJ/NRAO)

used up the subsidies anticipated for 2009 — and on 17 December the government postponed a round of awards planned for January until 1 March.

For a longer version of this story, see <http://tinyurl.com/9nxa5h>.

## US firms power up to build advanced batteries

Fourteen US companies have formed a coalition to make advanced lithium-ion batteries for vehicles. The National Alliance for Advanced Transportation Battery Cell Manufacture is proposing to build one or more large manufacturing plants that would be shared by its members.

Members of the alliance include Altair Nanotechnologies of Reno, Nevada, 3M of Saint Paul, Minnesota, and Johnson Controls-Saft, a joint American-French battery-production enterprise. The US Department of Energy's Argonne National Laboratory in Illinois will have an advisory role. The alliance is seeking government support for the majority of the venture, estimated to cost between \$1 billion and \$2 billion over five years.

## Health organization lays plans for major biobank

Health-care provider Kaiser Permanente based in Oakland, California, is building a biobank that could rival the world's largest in size and scope. The organization has received a grant of US\$8.6 million from the philanthropic Robert Wood Johnson Foundation in Princeton, New Jersey, to collect DNA specimens from 200,000 people enrolled in Kaiser's medical-care plan.

Kaiser will link the DNA to participants'

existing electronic medical records and to a planned database of environmental information, with the aim of reaching a total of 500,000 participants by 2012 — making it much larger than any other current biobank in the United States. The data will be available to researchers around the world for studying how genes and other factors influence disease risk and health.

For a longer version of this story, see <http://tinyurl.com/9bvr5d>.

## NASA awards contracts to supply space station

Orbital Sciences of Dulles, Virginia, and Space Exploration Technologies (SpaceX) of Hawthorne, California, have secured contracts from NASA to ferry cargo to the International Space Station (ISS) from 2010.

The contracts, which could each be worth up to US\$3.1 billion, will see the private carriers supply 40–70% of the space station's US cargo each year until the end of 2016. The firms will be stepping into the breach left when NASA retires its space shuttles in 2010.

In September 2008, SpaceX launched the first privately developed liquid-fuel rocket and its cargo into low Earth orbit. The company's ISS deliveries will probably be made with Falcon 9, a larger version of the rocket, which is expected to make its first launch within months.

### Correction

The News Feature 'Phoenix: a race against time' (*Nature* **456**, 690–695, 2008) suggested that Honeybee Robotics failed to follow revised designs that would have corrected a flaw in a part of the NASA Phoenix spacecraft. A subsequent investigation by mission managers at NASA's Jet Propulsion Laboratory found that Honeybee was not at fault. See Correspondence, page 26.



Subsidies may run short for Spain's solar parks.

# The new boss in town

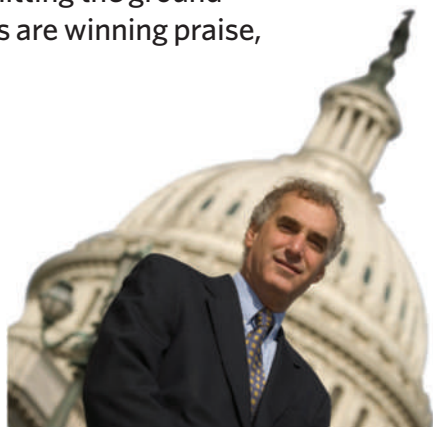
Barack Obama's transition team is hitting the ground running, and its speed and openness are winning praise, as **David Goldston** reports.

**T**he weekend after the US presidential election, I was in Chicago for a conference. At about 4 p.m. on Saturday afternoon, I stopped at a crowded coffee shop on one of the city's busiest streets and ran into two of president-elect Barack Obama's foreign-policy aides, who were working unobtrusively on their laptops at one of the small tables. "What are you doing here?" I asked, thinking that the first Saturday after a two-year campaign might be a day off. But they didn't comment about being at work. They said that the campaign office had now shut down and the transition office wasn't ready yet — so they had to work somewhere.

The story is emblematic of the Obama transition, which seems to have impressed just about everyone in Washington DC with its speed, diligence, expertise and openness. Although there have been reports of tensions at some agencies, particularly NASA, one usually partisan White House official told me that the Obama team gave him "this squishy good-government feeling I don't get very often". And someone who worked on the Clinton transition noted that in 1992 "we didn't even meet" until several weeks after election day.

Although some top Obama aides remain in Chicago, the bulk of the transition operation is housed in a nondescript government building in Washington. The setting is appropriate: although the level of excitement is high and the pace frenetic, the actual work is the day-to-day stuff of government — lots of meetings, writing papers, sending ideas up the chain of command with uncertain results, and racing to meet deadlines. "It seems like there's a deadline every day," one of the transition staff lamented to me. "They obviously can't all be meaningful."

There are two types of transition team: policy teams that are organized by topic and designed to look at big-picture issues, and review teams for each federal agency. The names of all the teams and their members are available on the transition website ([www.change.gov](http://www.change.gov)). In an unprecedented move, the site also posts all the written proposals that interest groups have submitted. The transparency is refreshing and has been lauded by public-interest groups, although it could limit the kinds of ideas the transition teams hear. Some groups may be reluctant to 'think out loud' or to offer ideas that diverge from their party line because everything



## PARTY OF ONE

will become public. But no one seems to be complaining. "We may take an extra day before sending something in" to make sure it's fully vetted, but the content is basically the same, said Tobin Smith of the Association of American Universities. Unsurprisingly, transparency does not extend to internal recommendations to Obama. The transition team has not been making its own papers public, and its members have been told not to speak to the media.

Although the transition team has been soliciting ideas on every imaginable issue, in early December, as economic news became ever more grim, the stimulus package that Obama hopes to see enacted in January began to dominate. Putting together multibillion-dollar proposals at high-speed is a little "disconcerting", one of the transition staff told me, noting that most of the team members are more used to worrying about budget constraints. Smith said Tom Kalil, a top transition official on science and innovation, called him on Monday 8 December and said he needed any ideas for a stimulus bill by that Thursday.

As the economy worsened, the criteria for what might be included in a bill apparently broadened. Smith said that earlier on, the teams seemed to be looking only for stimulus projects that could spend all their money in a year; by December that had stretched to three-year projects. "They realized that it can take projects a while to ramp up," said Smith, although they still want ideas that are "shovel-ready".

The transition team has been running its proposals past the career staff at the Office of Management and Budget (OMB), an OMB staffer told me. He said that the OMB has been asked to review technical matters such as

whether the numbers look realistic and how quickly the money could be spent. In policy areas in which the OMB and transition staff have a good and open relationship, discussions of the proposals can be extensive. The transition teams have eliminated or altered spending ideas because of OMB guidance, he said.

Before the stimulus temporarily shunted everything else aside, the transition teams had been meeting with just about every group under the sun on a wide range of policies. Francesca Grifo, who heads the Union of Concerned Scientists' work on scientific integrity, said she has participated in about five meetings with teams covering relevant agencies. "In some cases, we reached out to them," she said, "in others, they asked to meet with us." The Environmental Protection Agency transition team asked the group to set up a meeting of scientific and public-interest groups that had worked on scientific-integrity issues. Transition sources indicate that Obama is likely to issue an executive order on scientific integrity shortly after taking office. At all the meetings, Grifo said, the transition teams asked many questions, but were careful not to give away anything about their own thinking.

Smith noted that science groups are playing a much more active role in this transition than in past ones, partly because Obama's team has been open and partly because of concerns that have accumulated during the Bush administration. "We started being active during the campaigns," he said. "And for the first time, we recommended names" of people for science and education posts.

The teams have also been meeting with each federal agency. The transition team for the Office of Science and Technology Policy (OSTP) has met with all the political and career staff at the OSTP at least once, a White House official told me. The meetings have been candid and friendly, and have focused on decisions that will need to be made in the first 100 days, such as questions on the switch from analogue to digital television, the shutdown of the space-shuttle programme, and the Intergovernmental Panel on Climate Change. Classified matters have had to wait because the OSTP transition team does not yet have the necessary clearances.

The official said it was very helpful that some of the OSTP transition team had served in the Clinton administration and understood the White House operation. But still, he said, at a human level, the process left him with a "funny feeling. You see the transition team walking around with their badges in a building you think of as your place, and you realize it's not your place any more."

**David Goldston is a project director at the Bipartisan Policy Center in Washington DC. Reach him at [partyofonecolumn@gmail.com](mailto:partyofonecolumn@gmail.com).**



# NEW EYES, NEW SKIES

The next 40 years will see telescopes that far outstrip any ever seen before. **Jeff Kanipe** profiles four of them; illustrations by **Lynette Cook**.

The armillary and astrolabe are now seldom seen outside museums and antique shops; but the telescope, which joined them in the observatories of early modern Europe 400 years ago, is still at the centre of the astronomical world. In optical precision, in the wavelengths that are used and in their sheer size, they have changed almost beyond recognition (see page 28). After two centuries in which they left no records other than the users' sketches, and a century in which their visions were recorded on photographic plates, they have in the past decades become entirely electronic. And they are now stationed everywhere — oceans, deserts, mountain tops and all kinds of orbit. But the job is still the same: collecting and focusing whatever information the Universe sends our way.

Yet for all its glorious 400-year history, the astronomical telescope's best days may still be to come. Telescopes currently in development show an unprecedented degree of technical ambition as they seek to provide near-definitive answers to questions that, a generation or two ago, it was hard to even imagine investigating.

To answer these questions, the telescopes profiled here will often work in complementary ways. The infrared capabilities of the James Webb Space Telescope and the radio acuity of the Square Kilometre Array will



both be used to probe the Universe at the time of its own 'first light' — the birth of the first stars and galaxies. The radio array will map the large-scale structure of the Universe, elucidating the role in that structure of 'dark matter' and 'dark energy', as will studies of the faintest galaxies by the Large Synoptic Survey Telescope and European Extremely Large Telescope. That behemoth and the orbiting Webb will, in turn, complement each other in their attempts to characterize planets around other stars with unprecedented detail.

This quartet, for all its ambition and expense, does not exhaust the possibilities being explored and wished for. The Atacama Large Millimeter/Submillimeter Array will soon revolutionize astronomy at its chosen wavelengths. Other projects are planned throughout the electromagnetic spectrum and beyond into the new realms of gravitational waves and neutrinos. These instruments are all being designed with specific scientific challenges in mind. But at the same time, all concerned hope devoutly to discover something as strange and unlooked for as Galileo's mountains on the Moon — or spots on the face of the Sun.

**Jeff Kanipe is a science writer based in Maryland. Lynette Cook is an artist and illustrator based in California.**

See Editorial, page 7, and online at [www.nature.com/astro09](http://www.nature.com/astro09).

## The James Webb Space Telescope

Like the Hubble Space Telescope, to which it is in some ways the successor, the James Webb Space Telescope (JWST) will be the orbital flagship of its generation. But whereas the Hubble sees mainly in the visible and ultraviolet, JWST is optimized for the infrared. That means it can see things hidden from the Hubble and its like by dust, and peer into the high-red-shift epoch just after the Big Bang at objects indiscernible at visible wavelengths — such as the first stars.

Astronomers at the Space Telescope Science Institute in Baltimore, Maryland, started their first plans for a follow-on instrument in 1989 — a year before the Hubble itself was launched. It should finally make it to the launch pad 24 years later. Although its design and cost have changed a few times over the past two decades (see *Nature* 440, 140–143; 2006), its main mission remains simple and visionary — to study unseen aspects of every phase of the history of the Universe. To do so, the telescope will make use of several innovative technologies, such as ultra-light-

weight optical systems made from beryllium, extremely sensitive infrared detectors and a cryocooler that can maintain the mid-infrared detectors at a frosty 7 kelvin indefinitely.

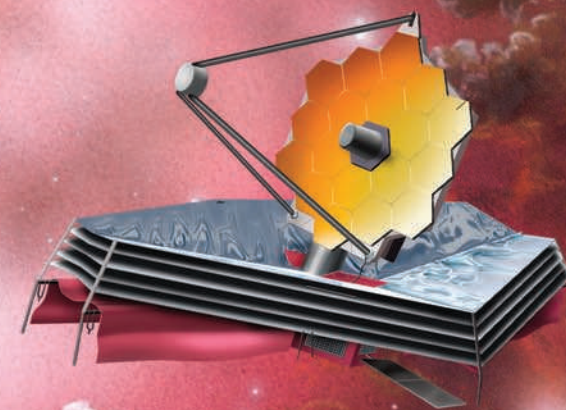
The most striking of the new technologies, though, affects the very heart of the telescope. JWST's designers wanted a mirror that would have been too large to fit into the payload fairing of any rocket available. So they designed one in segments, a mirror that could be launched folded up and then deployed to its full 6.5-metre diameter once the telescope settles into its final orbit, 1.5 million kilometres from Earth. That distance gives the telescope much more sky to look at than the Hubble gets, and keeps it cooler, too. But it has its downside: as yet there is no way to get there to fix any problems so, unlike, the Hubble, JWST has to work perfectly from the start.

At the moment, says John Mather, Nobel laureate and senior project scientist for JWST, the telescope is designed to last for at least five years, but longer may be possible. It will carry the ten years' worth of fuel, and the presence of the cryocooler means that, unlike earlier

infrared missions, its lifetime is not limited by a fixed supply of coolant. "If we are lucky and clever we hope to conserve fuel and perhaps run much longer," says Mather. "But we can't promise that." What Mather thinks he can promise is discovery. "We do not know which came first, black holes or galaxies, and we do not know how it happens that there is a massive black hole at the centre of almost every massive galaxy. If there are any surprises about the early Universe, I am guessing that they will be in these areas."

JWST is not just about deep space and distant epochs, though; it will also scrutinize the shrouded origins of objects closer to home — such as nascent solar systems, coalescing stars and star clusters amassing within dusty nebulae, says Matt Mountain, director of the Space Telescope Science Institute.

But where the telescope will really stand out will be its ability to probe the very early Universe. "JWST is so sensitive," says Mountain, "that we can take actual spectra of the very earliest objects you can just barely detect with Hubble."



### JWST

**Collection area:** 33 square metres

**First light:** 2013

**Cost:** US\$4.5 billion all in

**Unique selling point:** The best infrared possible



# A DEEP, MOVING IMAGE

## The Large Synoptic Survey Telescope

Sometimes telescopes see double not because of any aberration, but because that is the way the Universe works. The bending of light by intervening masses — called gravitational lensing — means that some galaxies are seen by Earthly observers in more than one place. By adding together survey image after survey image, and so measuring things that no individual image would show, the designers of the Large Synoptic Survey Telescope (LSST) hope to find a significant fraction of the 10,000 or so such images in every square degree of sky. They also hope to open up a neglected dimension in astronomy: time. As well as adding together images of the same part of space taken again and again to reveal new depth, they will compare those images to spot any differences, turning up a wealth of supernovae, asteroids and Kuiper belt objects on the fringe of the Solar System that would otherwise be missed. The telescope's proponents call it celestial cinematography.

The telescope will suck in celestial data by the terabyte every night, surveying almost all of the sky visible from Cerro Pachón, Chile, every week. Such coverage is made possible by an 8.4-metre primary mirror,

which will be ground so as to provide a field of view of 10 square degrees. That's 49 times the size of the full Moon, and more than 300 times the field of view of the Gemini telescopes, which have mirrors of similar size optimized for staring in a single spot. Over ten years, says Željko Ivezić, of the University of Washington in Seattle, the LSST system will look at everything in its field of view about 1,000 times. A massive amount of computing power will be used to correlate, compare and catalogue the torrent of data — and to make them all available

on the Internet. Anyone with a computer — students, and amateur and professional astronomers — will be able to participate in the process of scientific discovery.

Studies of objects that have been gravitationally lensed should reveal huge amounts about the structure of the Universe in general, and the distribution of dark matter and the effects of dark energy in particular. At the same time, though, LSST will mount a virtual space patrol, looking for potentially hazardous near-Earth asteroids. Astronomers already know where most of the big,

killing-off-species-wholesale asteroids are. LSST will be one of the tools that catalogues the vast majority of lesser asteroids still capable of smashing a city. But with a sensitivity to faint, transient events 1,000 times greater than ever previously achieved, the telescope will not restrict itself to the 'vermin of the skies' in Earth's backyard. It will observe vast distant cataclysms, such as collisions between neutron stars, and is all but sure of discovering whole new categories of transient events.

The project is overseen by the LSST Corporation, comprising more than 100 scientists and two dozen laboratories, universities and institutes based mainly in the United States. Although the project's design is still being worked out, the main mirror has already been cast. Astronomers with the corporation are hopeful that construction will begin as planned in 2011 and that first light will occur in 2015. In the subsequent ten-year survey, LSST will take stock of every object in the Universe, known, unknown and newly discovered. "For the first time in history," says Ivezić, "we will catalogue and study more celestial objects than there are people on Earth."

### Follow up around the world

LSST will be able to send out alerts on the Internet within a minute of seeing something untoward in the sky. The Las Cumbres Observatory Global Telescope will be eager to hear them. When completed, this will be a network of robotic telescopes distributed in two rings, one circling each hemisphere. Currently the network — a privately funded, non-profit organization — has just two telescopes up and running, one on Haleakala, Hawaii, and another in the Siding Spring Observatory in New South Wales, Australia. More are planned for sites in Mexico, the Canary Islands, Chile, South Africa and Australia. The final goal is to have two dozen or so 0.4-metre telescopes and a similar number of 1-metre telescopes. The smaller ones will be skewed towards educational use, the larger ones towards science, but there will be a lot of overlap. Once fully operational, the network will be able to keep a constant eye on new-found objects, such as supernovae and new asteroids, for days or weeks.



### LSST

**Collection area:** 35 square metres

**First light:** 2015

**Cost:** \$390 million to first light

**USP:** All of space, in real time



# SUPERSIZING THE HEAVENS

## The European Extra Large Telescope

Space-based observatories have much to recommend them; they see in wavelengths forever hidden from observers on Earth's surface, and with the crystal clarity of an airless sky. But they are hard put to compete in one of the most important aspects of telescope design: size. Whether you want a telescope for a backyard or a mountain top, the larger the light-gathering surface, the more photons you will catch. With a primary mirror measuring 42 metres across, the European Extremely Large Telescope (E-ELT) is

designed to pick out features only hinted at in the best images made by orbiting telescopes, detecting objects as faint as the planets around other stars. Its backers think that such a telescope could put satellites out of the optical-and-near-infrared business until similar sized behemoths can be assembled in orbit — or on the Moon.

Today E-ELT, a project of the European Southern Observatory (ESO), is halfway through a 3.5-year, €57.2-million (\$72.5-million) detailed design phase. Construction could start as soon as mid-2010.

It is not the only such giant planned: the Thirty Meter Telescope is being built by a public-private consortium including universities in North America, and the Giant Magellan Telescope is a collaboration between several US and Australian universities and research institutes (see *Nature* 452, 142–145; 2008). But E-ELT is the most audacious in scope and, with a budget of €950 million, the most expensive.

That said, the current E-ELT design is a scaled-down version. When originally discussed as the Overwhelmingly Large

Telescope (OWL), it was to have had a mirror 100 metres in diameter. The current 42 metres is more pragmatic. But if adaptive optics that constantly correct the mirror's shape to offset atmospheric turbulence work for a system of this size, as its designers hope, E-ELT should still have a spatial resolution 18 times better than that of the Hubble Space Telescope. Today's largest mirrors are those of the twin 10-metre Keck telescopes on Mauna Kea; E-ELT would have a light-gathering area almost nine times larger than both of them put together.

When it comes to thinking big, E-ELT's designers have the advantage, not shared by astronomers in the United States, of steady annual funding. ESO's management hopes to build the giant without asking the organization's 13 member states for more money, or having to hunt around for new partners. That could still change though: the telescope's optical design is not finalized, nor is its ultimate location. The Canary Islands, Morocco, Argentina and two sites in Chile are all under discussion.

All this is just background noise to Roberto Gilmozzi, E-ELT's principal investigator, who already has his eye on the telescope's potential discoveries. "One of the key goals of the telescope will be the quest for Earth-like planets around other stars," he says, "both by indirect methods — such as measuring radial velocity variations in the parent star — and by direct imaging and spectroscopy of their atmospheres." It will also aim to study proto-planetary systems around young stars in unprecedented detail.

"E-ELT will provide a 'quantum jump' in sensitivity over the present generation of telescopes — comparable to that of Galileo's telescope over the naked eye," says Gilmozzi, and that will affect astronomy at all scales. Galaxies that are now blurs will turn into fields of distinct stars. And on the largest of scales the telescopes will carry out a "physics experiment" that would not be possible with any other telescope, Gilmozzi says. By comparing ultra-precise velocity measurements of very distant objects taken several years apart E-ELT will make the first direct observations of how fast the Universe is expanding.

### E-ELT

Collection area: 1,385 square metres

First light: 2010s

Cost: \$1.4 billion to first light

USP: The clue is in the name

### Other ways to other planets

The international COROT mission, launched in 2006, has spent two years trying to detect dips in the brightness of stars caused by intervening planets. Later this year it will be joined by NASA's Kepler mission, and a follow-on European mission, PLATO, is planned for the middle of the next decade. All aim to discover Earth-sized and

smaller planets, particularly in or near the 'habitable zones' around stars where liquid water is possible on the surface. E-ELT might be used to confirm and follow up the findings of such missions — perhaps making spectroscopic measurements of a newly discovered planet's atmosphere, for instance. The combination of detection

from space with such satellites and study from Earth with a new generation of very large telescopes such as E-ELT means that astronomers will be able to start putting together profiles of other planetary systems, some of which may be similar to the Sun's, some of which will doubtless prove very different.



# LOOKING EVERYWHERE

## The Square Kilometre Array

It hasn't yet got a site; it hasn't yet got funding. But what the Square Kilometre Array (SKA) lacks in these pragmatic matters it more than makes up for in ambition: it will be the largest telescope of any kind in the world. Today's biggest single-aperture radio telescope, the Arecibo Observatory in Puerto Rico, has just 73,000 square metres of collecting area. SKA is aiming for a million. This collecting area will be distributed between thousands of small dishes, some grouped together in a central core, some trailing out along a vast crop-circle-like, multi-armed spiral to provide a total radius of a couple of thousand kilometres. That vast radius will give the system a resolution to match the incredible sensitivity provided by all that collecting area.

The result is a telescope that works over a frequency range of from 70 megahertz to 10 gigahertz, and boasts a huge field of view — 200 square degrees at frequencies of under 1 gigahertz, 1 square degree or so at higher frequencies. The telescope's final design will incorporate dishes for higher frequencies and flat 'phased-array' panels for the lower frequencies. An advantage of the phased-array approach is that observers can study different parts of the sky simultaneously. This design will allow the instrument to survey the sky 10,000 times faster than any radio telescope producing similar images. Research topics will cover more or less the whole reach of astronomy, from galaxy evolution and the cosmic jets spewed

out by quasars to the study of pulsars and supernova remnants.

The array is a truly international collaboration between institutions in 19 countries, including much of Europe, Argentina, Australia, Brazil, Canada, China, India, South Africa, Sweden and the United States. The consortium members hope to see it completed in 2020, and to keep producing science for 50 years. Taking into account the all-but-inevitable delay, later upgrades and the like, SKA could well last long enough to be the first great telescope of the twenty-second century.

According to James Cordes, chair of the US SKA Consortium, SKA will revolutionize understanding of the first galaxies, and of how the Universe developed its 'metals' (astronomers speak for all elements heavier than helium). Previous work has shown that just half a billion years after the Big Bang, galaxies were awash in carbon monoxide gas that could only have come from earlier stars. The array will be able to map that spread, and to measure CO molecules in the interstellar media of 'pregalactic objects' — the seeds of galaxies. But like everyone with a big telescope that they have fallen in love with, Cordes is also keen to stress the capacity for surprise inherent in the object of his affection. SKA, he says, will be a discovery machine. Who knows what it will see that lesser eyes on the sky have missed: "Perhaps extrasolar planets and flare stars of new types. Perhaps even exotica — such as evaporating black holes and extraterrestrial civilizations." ■

### SKA

**Collecting area:** Up to 1 million square metres  
**First light:** 2020s  
**Cost:** \$1.4 billion or more  
**USP:** All things radio, better

### Back to space

Earth-based telescopes find it relatively easy to grow to a giant size; in space things are more constrained. But astronomers at the Space Telescope Science Institute in Baltimore, Maryland, who are looking for the next big thing are thinking that it should be exactly that: big.

The Advanced Technology Large-Aperture Space (ATLAS) telescope is a proposed telescope that could only be launched on NASA's planned Ares V heavy-lift launch vehicle, the humans-to-the-Moon launcher that will be the largest rocket since the days of Apollo. ATLAS would have a larger mirror than the largest Earth-based telescopes today, with a diameter of perhaps 16 metres and a sensitivity some 40 times that of the Hubble telescope. It would provide yet another set of perspectives on dark matter and galactic evolution. Perhaps most exciting, though, would be its role in characterizing habitable planets — perhaps even inhabited ones — around other stars.



# CORRESPONDENCE

## Obama must match science rhetoric with action

SIR — I welcome the news of Barack Obama's smart choices for cabinet and sub-cabinet positions. This is a promising sign that the scientific process will once again be broadly valued here.

A group of graduate students recently asked my perspective on the impact of the current funding crunch. As a young investigator who began a tenure-track position in 2004, one year after the US National Institutes of Health (NIH) funding trajectory inflected from flat to negative, I understand the challenges to new scientists.

Since 2004, I have been the main author of three published articles. During this time, I have been awarded several small grants and have narrowly failed to attain an R01 application — the benchmark for promotion and tenure at most US universities. Meanwhile, I have watched junior faculty with similar records moved out of the tenure stream, seen scientific enthusiasm eroded and heard faculty members question the rewards of a career committed to research. What makes the crucial difference for me is tremendous support from my department and colleagues.

The former NIH director, Elias Zerhouni, strove to reverse the trend that has raised the average age of first funding from 37 in 1980 to 42 in 2007. As a result of his efforts, at least 1,650 investigators will receive their first R01 in 2009, up from 1,354 in 2006 (see *Nature* **456**, 153; 2008). I hope to be among them. My age? 39. My optimism? High, reflecting a supportive university environment.

Continued, active investment in new faculty not only ensures the success of young scientists but also directly influences the career decisions of PhD candidates. I hope our new president and administration truly recognize that our future is in science and technology, and that a

new generation of innovators needs their full support.

**Lisa Borghesi** Department of Immunology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, USA  
e-mail: borghesi@pitt.edu

## Problems with anti-plagiarism database

SIR — Sophisticated tools have been developed to detect duplicate publication and plagiarism, as noted in M. Errani and H. Garner's Commentary 'A tale of two citations' (*Nature* **451**, 397–399; 2008) and in your News story 'Entire-paper plagiarism caught by software' (*Nature* **455**, 715; 2008). To my surprise, one of these tools, Déjà vu (<http://spore.swmed.edu/dejavu>), classifies four of our publications as unverified duplicates. These report the analysis of Bruton's tyrosine kinase mutations associated with the rare disease X-linked agammaglobulinaemia (XLA) and of the database BTKbase.

Each of these is a genuinely different and independent report; they cover the development of the database and different analyses of the growing data set. The reason why they are branded as suspect cases is probably that the journal *Nucleic Acids Research*, in which three of them were published, has a special format for articles in their annual database issue.

Between 1995 and 2006, we published eight articles on BTKbase. The number of XLA cases recorded in the database has grown from 118 to 1,111 during this period. Several colleagues who maintain databases are also listed in Déjà vu. It is worrying that such legitimate articles written by research infrastructure developers and providers are labelled as unethical, just because of some overlap with previous papers as a result of a journal's strict formatting requirement.

Detection of fraud, including

duplications, is obviously crucial to the integrity of science. But it is unethical to list thousands of scientists in a public Internet service as suspects, without verifying the claims that are being made. Although the developers indicate that the data are provisional, there is still a risk that the listing will affect decisions on careers, promotions or research funding if individual cases are not investigated.

No professional scientist wants even the slightest suspicion of fraud to tarnish their scholarly reputation, so listed cases need to be closely investigated. To detect real duplicates, the full-length articles must be analysed, not just the abstracts — which occurred in the case of our publications.

**Mauno Vihinen** Institute of Medical Technology, FI-33014 University of Tampere, Finland  
e-mail: mauno.vihinen@uta.fi

## Honeybee and the Phoenix analysing instrument

SIR — In the News Feature 'Phoenix: a race against time' (*Nature* **456**, 690–695; 2008) you report on a problem that stopped the doors to the ovens on the Phoenix spacecraft's Thermal and Evolved Gas Analyzer (TEGA) instrument from opening fully. You note that the University of Arizona team responsible for TEGA noticed the door interference problem during engineering tests and sent revised designs to Honeybee Robotics of New York, but that Honeybee Robotics sent back new parts using the "original flawed designs".

We at the University of Arizona are concerned that Honeybee's responsibility for the TEGA door opening has been overstated. As the dust-cover manufacturer, the company played a very minor role in the overall TEGA construction. Most importantly, it was not responsible for establishing that the TEGA doors functioned properly after the dust-cover

assembly was installed.

Unfortunately, because the door-opening mechanism is a single-use actuator, which is buried deep within the instrument, we on the instrument team could not test it on the flight unit without a time-consuming process of dismantling and rebuilding the system.

In light of the publicity given to the door problem, we would like to set a few of the facts straight. When the door-opening problem was first observed during testing of an engineering prototype, we documented the needed change, along with some other changes, in an engineering drawing that was a modification to an earlier Honeybee drawing. Our drawing had more than a dozen changes on it, all but one of which were well documented with new dimensions called out. However, the change that related to the issue with the doors, although drawn properly, was not explicitly called out as a change in this way.

We should have asked to review the final drawing before metal was cut, but we did not. As we could not test the operation of the doors in the flight unit, we should have checked the parts when they arrived, but we did not. I recognize that my group at the University of Arizona operates a bit more informally than is current practice in the aerospace industry; this has served us well in the past in keeping costs down, but entails risk. We should have caught the problem and we didn't. In the end, the buck stops with us.

There is nothing associated with this event that changes our full confidence in Honeybee's competence. We have had an excellent working relationship with the company on this and several earlier projects, and Honeybee would clearly be our vendor of choice for similar projects in the future.

**William V. Boynton** TEGA instrument lead, Department of Planetary Sciences, The University of Arizona, Tucson, Arizona 85721, USA  
e-mail: wboynton@lpl.arizona.edu

Please see correction on page 16.



# CORRESPONDENCE

## Obama must match science rhetoric with action

SIR — I welcome the news of Barack Obama's smart choices for cabinet and sub-cabinet positions. This is a promising sign that the scientific process will once again be broadly valued here.

A group of graduate students recently asked my perspective on the impact of the current funding crunch. As a young investigator who began a tenure-track position in 2004, one year after the US National Institutes of Health (NIH) funding trajectory inflected from flat to negative, I understand the challenges to new scientists.

Since 2004, I have been the main author of three published articles. During this time, I have been awarded several small grants and have narrowly failed to attain an R01 application — the benchmark for promotion and tenure at most US universities. Meanwhile, I have watched junior faculty with similar records moved out of the tenure stream, seen scientific enthusiasm eroded and heard faculty members question the rewards of a career committed to research. What makes the crucial difference for me is tremendous support from my department and colleagues.

The former NIH director, Elias Zerhouni, strove to reverse the trend that has raised the average age of first funding from 37 in 1980 to 42 in 2007. As a result of his efforts, at least 1,650 investigators will receive their first R01 in 2009, up from 1,354 in 2006 (see *Nature* **456**, 153; 2008). I hope to be among them. My age? 39. My optimism? High, reflecting a supportive university environment.

Continued, active investment in new faculty not only ensures the success of young scientists but also directly influences the career decisions of PhD candidates. I hope our new president and administration truly recognize that our future is in science and technology, and that a

new generation of innovators needs their full support.

**Lisa Borghesi** Department of Immunology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, USA  
e-mail: borghesi@pitt.edu

## Problems with anti-plagiarism database

SIR — Sophisticated tools have been developed to detect duplicate publication and plagiarism, as noted in M. Errani and H. Garner's Commentary 'A tale of two citations' (*Nature* **451**, 397–399; 2008) and in your News story 'Entire-paper plagiarism caught by software' (*Nature* **455**, 715; 2008). To my surprise, one of these tools, Déjà vu (<http://spore.swmed.edu/dejavu>), classifies four of our publications as unverified duplicates. These report the analysis of Bruton's tyrosine kinase mutations associated with the rare disease X-linked agammaglobulinaemia (XLA) and of the database BTKbase.

Each of these is a genuinely different and independent report; they cover the development of the database and different analyses of the growing data set. The reason why they are branded as suspect cases is probably that the journal *Nucleic Acids Research*, in which three of them were published, has a special format for articles in their annual database issue.

Between 1995 and 2006, we published eight articles on BTKbase. The number of XLA cases recorded in the database has grown from 118 to 1,111 during this period. Several colleagues who maintain databases are also listed in Déjà vu. It is worrying that such legitimate articles written by research infrastructure developers and providers are labelled as unethical, just because of some overlap with previous papers as a result of a journal's strict formatting requirement.

Detection of fraud, including

duplications, is obviously crucial to the integrity of science. But it is unethical to list thousands of scientists in a public Internet service as suspects, without verifying the claims that are being made. Although the developers indicate that the data are provisional, there is still a risk that the listing will affect decisions on careers, promotions or research funding if individual cases are not investigated.

No professional scientist wants even the slightest suspicion of fraud to tarnish their scholarly reputation, so listed cases need to be closely investigated. To detect real duplicates, the full-length articles must be analysed, not just the abstracts — which occurred in the case of our publications.

**Mauno Vihinen** Institute of Medical Technology, FI-33014 University of Tampere, Finland  
e-mail: mauno.vihinen@uta.fi

## Honeybee and the Phoenix analysing instrument

SIR — In the News Feature 'Phoenix: a race against time' (*Nature* **456**, 690–695; 2008) you report on a problem that stopped the doors to the ovens on the Phoenix spacecraft's Thermal and Evolved Gas Analyzer (TEGA) instrument from opening fully. You note that the University of Arizona team responsible for TEGA noticed the door interference problem during engineering tests and sent revised designs to Honeybee Robotics of New York, but that Honeybee Robotics sent back new parts using the "original flawed designs".

We at the University of Arizona are concerned that Honeybee's responsibility for the TEGA door opening has been overstated. As the dust-cover manufacturer, the company played a very minor role in the overall TEGA construction. Most importantly, it was not responsible for establishing that the TEGA doors functioned properly after the dust-cover

assembly was installed.

Unfortunately, because the door-opening mechanism is a single-use actuator, which is buried deep within the instrument, we on the instrument team could not test it on the flight unit without a time-consuming process of dismantling and rebuilding the system.

In light of the publicity given to the door problem, we would like to set a few of the facts straight. When the door-opening problem was first observed during testing of an engineering prototype, we documented the needed change, along with some other changes, in an engineering drawing that was a modification to an earlier Honeybee drawing. Our drawing had more than a dozen changes on it, all but one of which were well documented with new dimensions called out. However, the change that related to the issue with the doors, although drawn properly, was not explicitly called out as a change in this way.

We should have asked to review the final drawing before metal was cut, but we did not. As we could not test the operation of the doors in the flight unit, we should have checked the parts when they arrived, but we did not. I recognize that my group at the University of Arizona operates a bit more informally than is current practice in the aerospace industry; this has served us well in the past in keeping costs down, but entails risk. We should have caught the problem and we didn't. In the end, the buck stops with us.

There is nothing associated with this event that changes our full confidence in Honeybee's competence. We have had an excellent working relationship with the company on this and several earlier projects, and Honeybee would clearly be our vendor of choice for similar projects in the future.

**William V. Boynton** TEGA instrument lead, Department of Planetary Sciences, The University of Arizona, Tucson, Arizona 85721, USA  
e-mail: wboynton@lpl.arizona.edu

Please see correction on page 16.

# CORRESPONDENCE

## Obama must match science rhetoric with action

SIR — I welcome the news of Barack Obama's smart choices for cabinet and sub-cabinet positions. This is a promising sign that the scientific process will once again be broadly valued here.

A group of graduate students recently asked my perspective on the impact of the current funding crunch. As a young investigator who began a tenure-track position in 2004, one year after the US National Institutes of Health (NIH) funding trajectory inflected from flat to negative, I understand the challenges to new scientists.

Since 2004, I have been the main author of three published articles. During this time, I have been awarded several small grants and have narrowly failed to attain an R01 application — the benchmark for promotion and tenure at most US universities. Meanwhile, I have watched junior faculty with similar records moved out of the tenure stream, seen scientific enthusiasm eroded and heard faculty members question the rewards of a career committed to research. What makes the crucial difference for me is tremendous support from my department and colleagues.

The former NIH director, Elias Zerhouni, strove to reverse the trend that has raised the average age of first funding from 37 in 1980 to 42 in 2007. As a result of his efforts, at least 1,650 investigators will receive their first R01 in 2009, up from 1,354 in 2006 (see *Nature* **456**, 153; 2008). I hope to be among them. My age? 39. My optimism? High, reflecting a supportive university environment.

Continued, active investment in new faculty not only ensures the success of young scientists but also directly influences the career decisions of PhD candidates. I hope our new president and administration truly recognize that our future is in science and technology, and that a

new generation of innovators needs their full support.

**Lisa Borghesi** Department of Immunology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, USA  
e-mail: borghesi@pitt.edu

## Problems with anti-plagiarism database

SIR — Sophisticated tools have been developed to detect duplicate publication and plagiarism, as noted in M. Errani and H. Garner's Commentary 'A tale of two citations' (*Nature* **451**, 397–399; 2008) and in your News story 'Entire-paper plagiarism caught by software' (*Nature* **455**, 715; 2008). To my surprise, one of these tools, Déjà vu (<http://spore.swmed.edu/dejavu>), classifies four of our publications as unverified duplicates. These report the analysis of Bruton's tyrosine kinase mutations associated with the rare disease X-linked agammaglobulinaemia (XLA) and of the database BTKbase.

Each of these is a genuinely different and independent report; they cover the development of the database and different analyses of the growing data set. The reason why they are branded as suspect cases is probably that the journal *Nucleic Acids Research*, in which three of them were published, has a special format for articles in their annual database issue.

Between 1995 and 2006, we published eight articles on BTKbase. The number of XLA cases recorded in the database has grown from 118 to 1,111 during this period. Several colleagues who maintain databases are also listed in Déjà vu. It is worrying that such legitimate articles written by research infrastructure developers and providers are labelled as unethical, just because of some overlap with previous papers as a result of a journal's strict formatting requirement.

Detection of fraud, including

duplications, is obviously crucial to the integrity of science. But it is unethical to list thousands of scientists in a public Internet service as suspects, without verifying the claims that are being made. Although the developers indicate that the data are provisional, there is still a risk that the listing will affect decisions on careers, promotions or research funding if individual cases are not investigated.

No professional scientist wants even the slightest suspicion of fraud to tarnish their scholarly reputation, so listed cases need to be closely investigated. To detect real duplicates, the full-length articles must be analysed, not just the abstracts — which occurred in the case of our publications.

**Mauno Vihinen** Institute of Medical Technology, FI-33014 University of Tampere, Finland  
e-mail: mauno.vihinen@uta.fi

## Honeybee and the Phoenix analysing instrument

SIR — In the News Feature 'Phoenix: a race against time' (*Nature* **456**, 690–695; 2008) you report on a problem that stopped the doors to the ovens on the Phoenix spacecraft's Thermal and Evolved Gas Analyzer (TEGA) instrument from opening fully. You note that the University of Arizona team responsible for TEGA noticed the door interference problem during engineering tests and sent revised designs to Honeybee Robotics of New York, but that Honeybee Robotics sent back new parts using the "original flawed designs".

We at the University of Arizona are concerned that Honeybee's responsibility for the TEGA door opening has been overstated. As the dust-cover manufacturer, the company played a very minor role in the overall TEGA construction. Most importantly, it was not responsible for establishing that the TEGA doors functioned properly after the dust-cover

assembly was installed.

Unfortunately, because the door-opening mechanism is a single-use actuator, which is buried deep within the instrument, we on the instrument team could not test it on the flight unit without a time-consuming process of dismantling and rebuilding the system.

In light of the publicity given to the door problem, we would like to set a few of the facts straight. When the door-opening problem was first observed during testing of an engineering prototype, we documented the needed change, along with some other changes, in an engineering drawing that was a modification to an earlier Honeybee drawing. Our drawing had more than a dozen changes on it, all but one of which were well documented with new dimensions called out. However, the change that related to the issue with the doors, although drawn properly, was not explicitly called out as a change in this way.

We should have asked to review the final drawing before metal was cut, but we did not. As we could not test the operation of the doors in the flight unit, we should have checked the parts when they arrived, but we did not. I recognize that my group at the University of Arizona operates a bit more informally than is current practice in the aerospace industry; this has served us well in the past in keeping costs down, but entails risk. We should have caught the problem and we didn't. In the end, the buck stops with us.

There is nothing associated with this event that changes our full confidence in Honeybee's competence. We have had an excellent working relationship with the company on this and several earlier projects, and Honeybee would clearly be our vendor of choice for similar projects in the future.

**William V. Boynton** TEGA instrument lead, Department of Planetary Sciences, The University of Arizona, Tucson, Arizona 85721, USA  
e-mail: wboynton@lpl.arizona.edu

Please see correction on page 16.



## COMMENTARY

## Time to turn off the lights

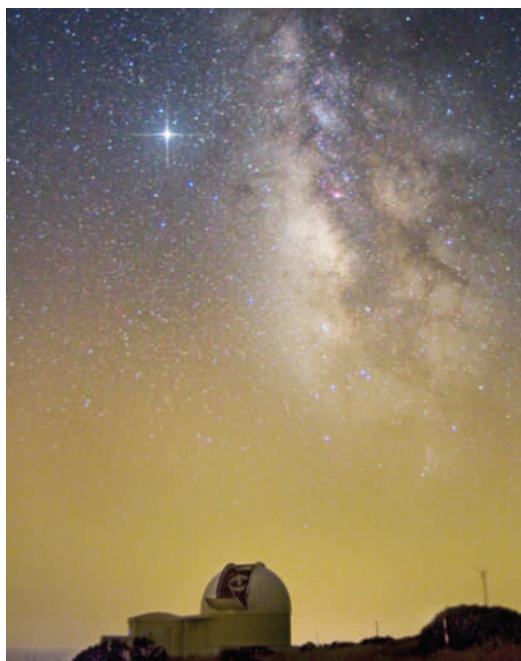
Cities needlessly shine billions of dollars directly into the sky each year and, as a result, a fifth of the world's population cannot see the Milky Way. **Malcolm Smith** explains why a dark sky has much to offer everyone.

Over the past six decades, professional and amateur astronomers have been pioneering efforts to curb light pollution to protect the viability of their observatories. During the 2009 International Year of Astronomy, particularly through the Dark Skies Awareness project<sup>1</sup>, astronomers can find allies in a common cause to convince authorities and the public that a dark sky is a valuable resource for everyone. Reducing the number of lights on at night could help conserve energy, protect wildlife and benefit human health.

The most persuasive arguments for lighting control are economic ones. Estimates by the International Dark-Sky Association, based on work from satellite images<sup>2</sup>, show that cities needlessly shine billions of dollars directly into the sky each year. As education on these issues improves, some cities are now realizing the benefits of controlling such energy waste through better-quality lighting, thereby reducing dangerous glare and confusing lighting clutter. Recent cost-saving measures in New York include lighting codes that require dimmers and motion-sensor-enabled light switches in many buildings.

Although humans are generally comfortable in artificially lit environments, man-made light sources can sow confusion for other animals, often with deadly results. For example, initiatives run by the National Audubon Society in New York, or Toronto's Fatal Light Awareness Program, focus on the effects that lights can have on bird migrations. In more and more cities in Canada and the United States, switching off at least some of the light in nearly empty skyscrapers reduces the unnecessary annual slaughter of millions of migrating birds in North America<sup>3</sup>. Some authorities now recognize that lights that are switched off during the migration season are probably not necessary. Yet, some cities still switch them back on when migration is over. There are many other examples and opportunities for cooperation<sup>3-5</sup>.

Humans are affected too, especially as urban areas expand worldwide<sup>6</sup>. More than two-thirds of the US population and one-fifth of the world's population have lost naked-eye visibility of the Milky Way<sup>7</sup>. This has a subtle cultural impact. Without a direct view of the stars, mankind is cut off from most of the



Village lights reflect off Saharan dust beyond the Teide Observatory on Tenerife, obscuring stars near the horizon.

Universe, deprived of any direct sense of its huge scale and our tiny place within it. Satellite and nightscape photos have illustrated beautifully just how much humans stand to lose. The work of The World at Night photographers has been incorporated into efforts for the International Year of Astronomy<sup>8</sup>. The International Astronomical Union has recently signed a memorandum of understanding with the United Nations Educational, Scientific and



Cultural Organization (UNESCO) to pursue the goal of identifying and protecting astronomical sites of historic significance. And the UNESCO-supported Starlight Initiative<sup>9</sup> has a major international programme investigating more deeply the issues discussed here. The potential for profitable ecotourism, associated with protected, natural, starlit skies is now being understood in many countries. The US National Park Service has launched a highly successful night-time-visits service in a joint effort between astronomers and park staff.

Even our physiology is deeply entwined with dark-light cycles. Recent research shows that light at night triggers signals that cause a reduction in the normal production of melatonin, a

suppressant of cell division in cancerous tissues<sup>10</sup>. It is imprudent to draw any grand conclusions yet, but we should continue to work with medical professionals to see if there are opportunities for cooperation.

Our relationship with artificial light is complicated and changing. Humans innately fear the darkness, and modern society relies on light as a security measure even though there is no evidence that controlling light wastage increases crime levels. Moreover, blinding automobile headlights require higher levels of street lighting to recover visibility of what's behind them. In many places in the world, one can drive around moderately lit, smaller towns with side lights. In the countryside, moderate headlights can be supported with catseye road reflectors.

As light fixtures with energy-efficient, light-emitting diodes start to become more widespread, the International Dark-Sky Association has been harnessing the support of professional lighting engineers and other non-astronomy interests to encourage responsible, innovative lighting and guidance design. Although it will be challenging, we can work together to reduce the form of pollution and energy waste that is, arguably, the most easily mitigated. ■

**Malcolm Smith** is an astronomer at the Cerro Tololo Inter-American Observatory, Casilla 603, La Serena, Chile.

e-mail: msmith@ctio.noao.edu

1. [www.astronomy2009.org/globalprojects/cornerstones/darkskiesawareness/](http://www.astronomy2009.org/globalprojects/cornerstones/darkskiesawareness/)
2. Isobe, S. I. & Hamamura, S. *Astrophys. Space Sci.* **273**, 289-294 (2000).
3. Deda, P., Elbertzagen, I. & Klusmann, M. in *Starlight: A Common Heritage. Proceedings Starlight Conference 2007* (eds Marin, C. & Jafari, J.) 177-183 (2007). Available at [www.starlight2007.net/proceedings.htm](http://www.starlight2007.net/proceedings.htm)
4. Longcore, T. & Rich, C. *Front. Ecol. Environ.* **2**, 191-198 (2004).
5. Rich, C. & Longcore, T. (eds) *Ecological Consequences of Artificial Night Lighting* (Island Press, 2006).
6. *State of World Population 2007: Unleashing the Potential of Urban Growth* (United Nations Population Fund, 2007). Available at [www.unfpa.org/swp/2007/english/introduction.html](http://www.unfpa.org/swp/2007/english/introduction.html)
7. Cinzano, P., Falchi, F. & Elvidge, C. D. *Mon. Not. R. Astron. Soc.* **328**, 689-707 (2001).
8. [www.astronomy2009.org/globalprojects/specialprojects/worldatnight/](http://www.astronomy2009.org/globalprojects/specialprojects/worldatnight/)
9. Marin, C. & Jafari, J. (eds) *Starlight: A Common Heritage. Proceedings Starlight Conference 2007* (2007). Available at [www.starlight2007.net/proceedings.htm](http://www.starlight2007.net/proceedings.htm)
10. Navara, K. J. & Nelson, R. J. *J. Pineal Res.* **43**, 215-224 (2007).

See Editorial, page 7, and [www.nature.com/astro09](http://www.nature.com/astro09).

## ESSAY

## Mankind's place in the Universe

Technological developments in astronomy have long helped to answer some of the greatest questions tackled by humanity, recounts **Owen Gingerich**.

Just over 400 years ago, news swept across Europe of an astonishing device — two lenses within a tube — that almost magically brought distant views closer. Its exact origins are lost in a mist of controversy, but it is known that one Hans Lipperhey in Middelburg, the Netherlands, applied for a patent in October of 1608. His application was turned down on the grounds that the unnamed novelty was already common knowledge.

Seventeen months later a small booklet by an Italian mathematics professor announced wonderful discoveries made with this newly devised spyglass. The title page may have implied that the 'perspicillum' was Galileo Galilei's own invention, but he should be forgiven any exaggeration, for what Galileo had done was to turn a popular carnival toy into a scientific instrument. In December 1609 he used his spyglass to discern craters, mountains and plains on the Moon. The 400th anniversary of these astonishing observations sets the stage for what the United Nations has designated the International Year of Astronomy.

By January 1610, Galileo had further found four bright moons of the planet Jupiter (it would be his contemporary, Johannes Kepler, who named them 'satellites'). Before that year was out Galileo had also found the phases of Venus, proving that it revolved about the Sun.

The implications of such findings were immense. Although Nicolaus Copernicus's *De revolutionibus orbium coelestium* ('On the Revolutions of the Heavenly Spheres'; 1543) — the work that proposed a Sun-centred or 'heliocentric' system — was already nearly seven decades old, it was taken only as a manual for computing positions of the planets; by no means was it thought to be a physically real description of the cosmos. How ridiculous to think of living on a globe that spun around once every 24 hours! And how could Earth keep the Moon in tow if it whirled around the Sun every year?

It was Galileo's work that made it intellectually respectable to believe in heliocentrism as a physical reality. Galileo saw that the Moon was



Earth-like, rather than a pure ball of crystalline aether: the perceived distinction between Earth, home of corruption and decay, and the eternal, aethereal heavens was crumbling. Jupiter had no trouble keeping its moons in tow, and everyone agreed that Jupiter was moving. Why could a moving Earth not keep its Moon on an invisible tether? Galileo's skillful polemics on this matter became full-blown in his 1632 *Dialogue on the Two Chief World Systems* — a work that failed to conclusively prove the

in his *A Perfit Description* (1576), placed a "habitable [home] for the elect" among the stars.

It is not surprising that clerical authorities for the most part resisted the new cosmology. More than this, the Roman Catholic authorities were furious that Galileo, a mere amateur theologian, had the audacity to tell them how to interpret scripture. It was clearly a turf battle. Galileo was brought to trial in Rome, allowed to plead innocent to charges of vehement suspicion of heresy, but convicted of disobeying orders (in

teaching the Copernican cosmology), and sentenced to house arrest for the remainder of his life.

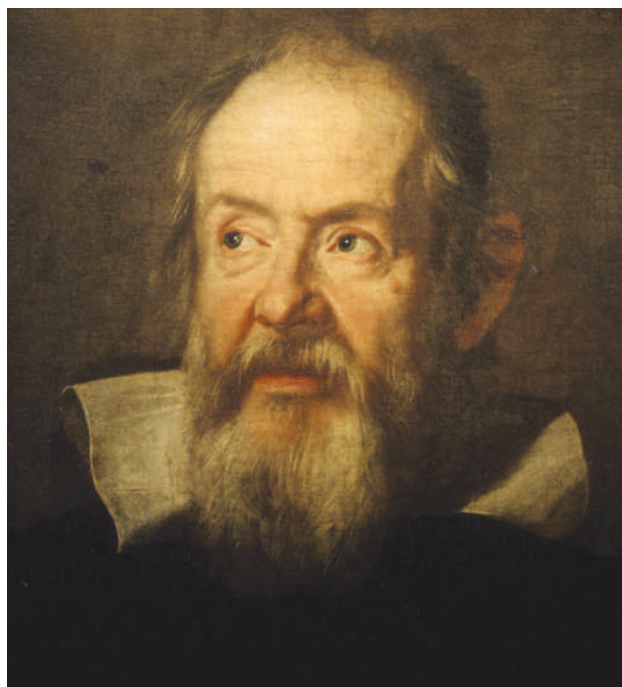
"New philosophy calls all in doubt," lamented the poet John Donne. "'Tis all in pieces, all coherence gone."

Galileo's persuasiveness was helped along by Kepler, whose landmark volume of theoretical astronomy was also published in 1609. His aptly named *Astronomia nova* or 'New Astronomy' relied heavily on looking at the physical causes of planetary motion — a critical departure from the past, when astronomers used strictly geometrical modelling to explicate the heavens. Even Kepler's teacher and mentor, Michael Maestlin, urged him to forget about physics and stick to astronomy (that is, geometry). But Kepler believed in a physically real cosmos, and even ahead of Galileo advocated the Sun-centred system.

Heliocentrism as it had come down from Copernicus, however, was flawed. Copernicus used entirely different geometrical models for the lati-

tudes of planets than for the longitudes, whereas for a physically real system, the same model should work for both coordinates. According to Copernicus, the planets moved in circles (the main ones eccentrically placed with respect to the Sun), and Earth did not move faster when closest to the Sun. It was Kepler's requirement for plausible physical explanations that drove him ultimately to postulate an ellipse as the basic form of planetary orbits, ironing out these difficulties.

Alongside Galileo's telescopic observations, Kepler's improvements in planetary theory eventually brought about general acceptance of the Copernican cosmology. In the introduction to *Astronomia nova*, Kepler wrote:



Galileo's observations of the Moon had far-reaching consequences.

motion of Earth, yet was persuasive enough to land him in deep trouble with the Inquisition.

### Heavenly vision

Aristotle had, some two millennia earlier, described a mundane world of earth, air, fire and water that was separate from the eternal purity of the aethereal spheres in the heavens. The ecclesiastical vision of a heavy, central Earth upon which the human drama played fit nicely into this Aristotelian cosmos — volcanoes belching out smoke and flames hinted at the hellfire below, whereas the realm of God and the elect lay beyond the starry firmament. Even Thomas Digges, the English astronomer who supported the "noble Copernican philosophy"



WELLCOME LIB., LONDON

"Perhaps there is someone whose faith is too weak to believe Copernicus without offending his piety. Let him stay at home and mind his own business. Let him assure himself that he is serving God no less than the astronomer to whom God has granted the privilege of seeing more clearly with the eyes of the mind."

Before and since, astronomical devices have been at the heart of mankind's literal search for our place in the Universe. It is not an inexpensive endeavour. Tycho Brahe boasted that his observations (which did not involve a telescope but rather carefully graduated naked-eye sightings using quadrants and other devices) cost the King of Denmark more than a ton of gold — a huge fortune in the sixteenth century. Today, telescopes can cost billions of dollars. It could be argued that such devices are ultimately as useful (or useless) to human progress as poetry. But they are funded because, like the greatest poetry or the building of a Gothic cathedral, cosmology is a voyage of the human spirit, and as such, a point of personal or national pride. It is worth a great price to understand our place in the intricate web of time and space.

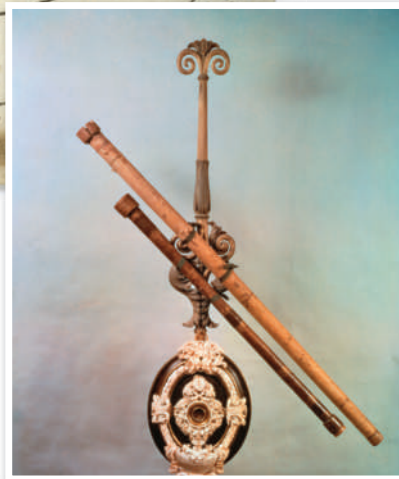
### The farthest reaches

After Galileo's work and Kepler's refinements — replacing a convex objective lens and concave eyepiece with two convex lenses, allowing for a wider field of view and accommodating crosshairs — the next major leap forward for observational astronomy would not come for nearly two centuries. Towards the end of the eighteenth century, William Herschel, an inspired musician-turned-amateur-astronomer, began his seemingly quixotic quest to determine the three-dimensional structure of the Universe. In the process, he discovered a new planet — Uranus — an event that changed his status from amateur to professional. The most inventive astronomer of his age, Herschel set on a path to ever-greater light-gathering power by building huge reflecting telescopes. He ended up counting thousands of stars and discovering hundreds of clusters and nebulae — the gateway to the vast outer Universe beyond the traditional sphere of naked-eye stars.

A century later, thanks to engineering advances, the discovery of how to silver glass



**Fresh views:**  
Thomas Digges's  
heliocentric Universe  
(above) and Galileo's  
perspicillum.



mirrors, and the ability of photography to capture images of stars and nebulae, astronomers finally marched through that portal. In 1859, in a chemical laboratory in Heidelberg, chemist Robert Bunsen and physicist Gustav Kirchhoff discovered the key to unlocking the chemical secrets of the Universe via the distinctive spectral lines associated with individual elements. Meanwhile, rich philanthropists poured cash into instruments of discovery, and from the new giant telescopes emerged a sequence of consciousness-raising discoveries. Among them was the immense size of the Milky Way, far beyond anything dreamed of by Herschel. Then came the realization that the Milky Way was but one of myriad galaxies stretching beyond by millions (and even billions) of light years, rushing away from each other as if from a cataclysmic explosion billions of years ago.

This vast increase in the size and age of the perceived cosmos set the stage for an angst as deep as that caused by the displacement of the ancient human-centred cosmology:

**"Will we soon answer the next psyche-shifting question: are we alone?"**

what significance do rational, observing mortals have in the wilderness of a near-infinite space-time continuum? This existential question has been simmering for decades and undoubtedly drives our willingness to invest taxpayers' money in further cosmic explorations.

Today, the electronic revolution has provided charge-coupled devices to replace photographic plates that have nearly 50 times their efficiency; has made feasible the collection and handling of data in the trillions of bits; and has enabled high-speed adaptive optics to compensate for atmospheric variations, allowing astronomers to take full advantage of large-aperture

Earth-based telescopes. The stellar and galactic denizens of the Universe can now be catalogued as never before. That has included, in recent years, the discovery and cataloguing of exoplanets — those beyond our Solar System.

NASA's Kepler mission, scheduled for launch in March 2009, is the first craft designed explicitly to look for Earth-like exoplanets. Don't expect radio communications from intel-

ligent life to come from one of those planets. But other exciting signals could surely be found. Today, spectroscopy can be used to spy on the atmospheres of far-away planets, and look for chemical traces that may hint at signs of life — even an extremely basic form of life could produce an oxygen signal in a planet's atmosphere. Such studies may help in answering the next psyche-shifting question: are we alone?

The International Year of Astronomy might well launch the next intellectual revolution in our understanding of our place in the Universe. Could this have as much of an impact on society as Galileo and Kepler's entrenchment of the heliocentric view? Only time will tell.

**Owen Gingerich** is professor emeritus of astronomy and history of science at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, and author of *God's Universe*. e-mail: ginger@cfa.harvard.edu

See <http://tinyurl.com/57589t> for further reading. See also [www.nature.com/astro09](http://www.nature.com/astro09).

AKG-IMAGES

## BOOKS &amp; ARTS

## Visions of ourselves

The view of our planet from space is beautiful and humbling, yet this shift in human perspective has not altered how we care for our environment, argues **Charles Cockell**.

**Earthrise: How Man First Saw the Earth**

by Robert Poole

Yale University Press: 2008. 236 pp. \$26

It is human nature to try to understand our place in the cosmos. Changes in our perception of human significance have accompanied many revolutions in scientific thought, such as the theory of evolution and establishing the order of the planets in the Solar System. So it is perhaps unsurprising that as soon as we developed the ability to launch projectiles into space, one of the first things we did was to use them to look back at ourselves.

The history and significance of the view of Earth from space is the subject of Robert Poole's *Earthrise*. Although its publication is titled and timed to mark the 40th anniversary of the first snapshots of Earth rising over the Moon's horizon on 24 December 1968, the book also traces the long journey that led to *Apollo 17*'s iconic image, 'Blue Marble', captured on 7 December 1972.

Of the many attempts before the space age to understand how Earth looks from afar, the Roman historian Seneca provides the most remarkable quote: "Is this that pinpoint which is divided by fire and sword among so many nations? How ridiculous are the boundaries set by mortals." The first human view of the curvature of Earth is attributed to the high-altitude balloon flight of *Explorer II* in 1935. Floating 22 kilometres above the ground, two intrepid aeronauts inside the balloon's pressurized gondola snapped a photograph that later appeared in *National Geographic*, with a line drawn across the horizon to illustrate that the curvature had been seen.

A robot took the first picture of the whole Earth in one frame. Black and white photographs of the planet, developed on board the spacecraft and scanned by a light that picked up different shades of grey, were beamed back from Lunar Orbiter 1 on 23 August 1966.

NASA photographer Richard Underwood emerges in the book as an early crusader for photographs of Earth from space taken by humans. He battled against bureaucrats who either thought it was a waste of astronaut time or could not see the scientific value. But during *Apollo 8*'s mission, on Christmas Eve 1968, *Earthrise* over the Moon was captured for the first



time. Even though the astronauts were trained for this eventuality, their transcripts reveal that they found the beauty of the sight profound and unexpected. The view of Earth from space has deeply affected many astronauts, particularly the lunar-module pilots who, while their fellow countrymen were wandering about on the Moon, had the time not merely to stare at Earth, but also to think about it.

In my view, however, the sociological significance of the *Earthrise* photograph is exaggerated — a possibility that Poole might have explored more. He devotes just a page to political theorist Hannah Arendt's thoughts and author Kurt Vonnegut's contrarian observation about the Earth images: "It looks so clean. You can't see all the hungry, angry earthlings down there — and the smoke and the sewage and trash and sophisticated weaponry." And yet, the idea that the environmental movement was caused by this view of Earth is controversial. One might equally say that our growing understanding of the connectedness of the biosphere was caused by the awakening in the 1960s to the global consequences of the use of atomic weapons and the effects of DDT insecticide. Yet the *Apollo* images were co-opted by the environmental movement and vocalized as another space-related analogy, 'Spaceship Earth', a concept that highlighted the planet's limited resources.

The promise of the future effect of this holistic view of Earth and its inhabitants may also be utopian. The revelations of Charles Darwin and Nicolaus Copernicus were not accompanied by a reduction in the number of wars as a result of a new humility. One might even make the case that understanding more fully the beautiful, but nevertheless insignificant scale of our planet might eventually dull us to acts of large-scale destruction as our field of play expands beyond the borders and histories of nations to territorial views of whole worlds.

Whatever one's opinion, seeing the view of our planet from space has been a psychological transition point. Like the first time one hears one's own recorded voice, or sees oneself on a video, we have been confronted with an objective view of ourselves that does not quite equate with our previous self-perception. *Earthrise* is a fascinating contribution to a growing discussion about how space exploration and settlement will change civilization. ■

**Charles Cockell** is professor of geomicrobiology at the Open University, Milton Keynes MK7 6AA, UK, and author of *Space on Earth: Saving Our World by Seeking Others*.

e-mail: c.s.cockell@open.ac.uk

See Editorial, page 7, and [www.nature.com/astro09](http://www.nature.com/astro09).



## Q&A: One giant leap for art

Astronaut **Alan Bean** stepped down onto the lunar surface during the 1969 *Apollo 12* mission, but left NASA in 1981 to devote himself to painting. With exhibitions of his work taking place this year to celebrate the 40th anniversary of the first Moon landing, he tells *Nature* how he attempts to convey his lunar experience.

### How did you become a painter?

Accidentally, I ended up becoming the first artist in history to go anywhere else beyond Earth and paint it. When I got ready to leave NASA, half of the astronauts wondered if this was a worthwhile thing to do with my life. And the others weren't sure. Now they all like it. They can see that I'm preserving one of the great adventures of humankind in a way that is meaningful for them.

### Why are your Moon paintings so colourful?

The Moon isn't colourful — it's grey. The sky is black and we run around in white suits. My first paintings were mostly grey, because that's what the Moon is. Now I realize the artist's job is not to reproduce reality.



### Why do you use Moon dust and tools from your mission?

The Moon was pretty rugged, so I wanted to put some texture in. The first texture I added was with normal art tools. Then I said "Why am I doing this? I've got Moon tools" — I've got the hammer I had on the Moon; I've got Moon boots.

A friend in charge of a museum requested the *Apollo 12* command module, and they shipped it to him in this huge box. He looked in the box and a lot of the charred heat shield had come off, so he vacuumed it up and sent it to me. I put that in there. Then



Alan Bean's *Rock 'n' Roll on the Ocean of Storms* shows the trials of collecting Moon rock.

one day I was looking at the patches from my suit and saying "Boy, those things are dirty". And it dawned

on me that they're dirty with Moon dust from the Ocean of Storms. I started cutting them up and put them in too.

### What are you working on now?

My current painting tells the story of when I fell down on the Moon and Pete Conrad (*Apollo 12*'s commander) came over and lifted me up with one finger. I'm lying on my butt leaning back and I've got my left arm out.

Pete's got his right forefinger around my left forefinger and he's lifting me up. I'm thinking of titling it *He Ain't Heavy, He's My Brother*.

### What is your view of the Moon landings 40 years on?

We went to the Moon, truthfully, in peace for all mankind. We were racing the Russians but we didn't try to do anything more than prove that we could do it. It raised the human consciousness on Earth because people said "Look what we can do". We've got to keep doing these things. ■

Interview by **Daniel Cressey**, a reporter for *Nature*.

## Voyaging to discovery, alone

**The Age of Wonder:  
How the Romantic Generation Discovered  
the Beauty and Terror of Science**  
by Richard Holmes

HarperPress/Pantheon: 2008/2009.  
380 pp/576 pp. £25/\$40

Everyone knows how to be a great scientist. First, you have to be really smart. Having awed your school science teacher is good. Humbling all the others at university is even better. Then you need to find a place where you can let your solitary genius come out. You might escape to

a coffee shop to think; or if your parents are really rich, to a trendy loft apartment or, better still, a remote and windswept cottage. There you will engage in the next and most crucial of stages: the creative torture.

There is an art to this. It is important your torment doesn't end too quickly — otherwise you'd show you had been working on a problem that was too simple. Yet if the torment never ends, you'll have nothing to tell anyone about and will remain unknown. After some time, ideally a few months, you get to have a 'eureka moment'. At that point all you have

to do is write up your discovery, accept the plaudits, be it from department colleagues or the Swedish Academy, and then if you can bear it, repeat from the start.

We smile at this recap but the basic vision — science as an endeavour of individual creative genius that explodes in an instant of discovery — is one we take for granted. Yet, as Richard Holmes describes in his new book *The Age of Wonder*, it was not always so. Most founders of modern science in the 1600s, such as Isaac Newton, rarely saw their work this way. For them the process was clinical, building on a slow accumulation of insight.

The big shift took place in the decades around 1800, a period called the Romantic era.



## Q&A: One giant leap for art

Astronaut **Alan Bean** stepped down onto the lunar surface during the 1969 *Apollo 12* mission, but left NASA in 1981 to devote himself to painting. With exhibitions of his work taking place this year to celebrate the 40th anniversary of the first Moon landing, he tells *Nature* how he attempts to convey his lunar experience.

### How did you become a painter?

Accidentally, I ended up becoming the first artist in history to go anywhere else beyond Earth and paint it. When I got ready to leave NASA, half of the astronauts wondered if this was a worthwhile thing to do with my life. And the others weren't sure. Now they all like it. They can see that I'm preserving one of the great adventures of humankind in a way that is meaningful for them.

### Why are your Moon paintings so colourful?

The Moon isn't colourful — it's grey. The sky is black and we run around in white suits. My first paintings were mostly grey, because that's what the Moon is. Now I realize the artist's job is not to reproduce reality.



### Why do you use Moon dust and tools from your mission?

The Moon was pretty rugged, so I wanted to put some texture in. The first texture I added was with normal art tools. Then I said "Why am I doing this? I've got Moon tools" — I've got the hammer I had on the Moon; I've got Moon boots.

A friend in charge of a museum requested the *Apollo 12* command module, and they shipped it to him in this huge box. He looked in the box and a lot of the charred heat shield had come off, so he vacuumed it up and sent it to me. I put that in there. Then



Alan Bean's *Rock 'n' Roll on the Ocean of Storms* shows the trials of collecting Moon rock.

one day I was looking at the patches from my suit and saying "Boy, those things are dirty". And it dawned

on me that they're dirty with Moon dust from the Ocean of Storms. I started cutting them up and put them in too.

### What are you working on now?

My current painting tells the story of when I fell down on the Moon and Pete Conrad (*Apollo 12*'s commander) came over and lifted me up with one finger. I'm lying on my butt leaning back and I've got my left arm out.

Pete's got his right forefinger around my left forefinger and he's lifting me up. I'm thinking of titling it *He Ain't Heavy, He's My Brother*.

### What is your view of the Moon landings 40 years on?

We went to the Moon, truthfully, in peace for all mankind. We were racing the Russians but we didn't try to do anything more than prove that we could do it. It raised the human consciousness on Earth because people said "Look what we can do". We've got to keep doing these things. ■

Interview by **Daniel Cressey**, a reporter for *Nature*.

## Voyaging to discovery, alone

**The Age of Wonder:  
How the Romantic Generation Discovered  
the Beauty and Terror of Science**  
by Richard Holmes

HarperPress/Pantheon: 2008/2009.  
380 pp/576 pp. £25/\$40

Everyone knows how to be a great scientist. First, you have to be really smart. Having awed your school science teacher is good. Humbling all the others at university is even better. Then you need to find a place where you can let your solitary genius come out. You might escape to

a coffee shop to think; or if your parents are really rich, to a trendy loft apartment or, better still, a remote and windswept cottage. There you will engage in the next and most crucial of stages: the creative torture.

There is an art to this. It is important your torment doesn't end too quickly — otherwise you'd show you had been working on a problem that was too simple. Yet if the torment never ends, you'll have nothing to tell anyone about and will remain unknown. After some time, ideally a few months, you get to have a 'eureka moment'. At that point all you have

to do is write up your discovery, accept the plaudits, be it from department colleagues or the Swedish Academy, and then if you can bear it, repeat from the start.

We smile at this recap but the basic vision — science as an endeavour of individual creative genius that explodes in an instant of discovery — is one we take for granted. Yet, as Richard Holmes describes in his new book *The Age of Wonder*, it was not always so. Most founders of modern science in the 1600s, such as Isaac Newton, rarely saw their work this way. For them the process was clinical, building on a slow accumulation of insight.

The big shift took place in the decades around 1800, a period called the Romantic era.



R. GENDLER/NASA

## Star poetry

**The Very Small Baseline Group Convenes at The Cat and Fiddle**  
by Neil Rollinson

A groaning table of empties makes up our Very Small Array — a barley-scented interferometer. Here we can study the cosmos and drink. We tune in to the microwave sky: to the froth at the edge of the universe. We sup in the dusk, everything glows with its own light: the hedgerow, lawn, the atoms inside the glass. The Milky Way sings in a half-inch of Guinness a song of distant past when the world was a moment old. We gather it all in our mugs, in a pub garden on the edge of the moors, looking down on Jodrell Bank: Queen of the red-light district, cocking her huge lug to the mayhem beyond our patch. The bats are in on it, hunting in ultrasound, catching moths in their fangs, while frogs bark in the meadows, one to the other, a vast unfathomable love-song. I finish my pint and add my glass to the phalanx: the more we drink the clearer we see, as any old soak will tell you. I tip back my head to look at the Pleiades and tumble, arse over tit, into the damp grass. I lie in my cups under the bling of the northern sky. I can hear it now, I can see it all clearly, all and nothing, just the whole sky blazing.

Poets paired with astronomers have created 16 new poems, including this one, that form part of a collection of 100 poems about the wonder of the Universe. ■

### Dark Matter: Poems of Space

Edited by Maurice Riordan and Jocelyn Bell Burnell

Calouste Gulbenkian Foundation: 2008. 208 pp. £8.50

The pursuit of progress by thinkers and artists became more a wait for a divine spark of inspiration than the steady toil of uncovering that had been accepted by their predecessors. And any means by which that spark could be nurtured was embraced.

To guide readers through the science and culture of this period, Holmes masterfully dips in and out of the life of Joseph Banks. He is an inspired choice. As a curly haired 26-year-old, Banks was aboard HM Bark *Endeavour* on the momentous day in April 1769 when it first glided into sight of Tahiti. Ostensibly Banks was just the expedition's plant collector, but he soon became more fascinated by the island's human inhabitants.

For most of the British crew — young men who had been away from female company for long months — the explorations in Tahiti were of one sort only, with the initial going rate being one ship's nail for one sexual encounter. That rate soon changed — as Holmes describes with gentle skill, the Tahitians were quick to grasp the workings of the market economy, and had a keen eye for the other useful metal objects aboard the ship. Hyperinflation set in, and at one point “there was a crisis when one of the *Endeavour's* crew stole a hundredweight bag of nails, and refused to reveal its whereabouts even after a flogging”.

Banks looked further. He took up Tahitian mistresses too, but systematically recorded the local language, studied their religious systems, and even hinted at the true functional significance of native actions that, at first, seemed to be merely bizarre. Within a few months he had helped set the stage for the modern science of anthropology.

Back in London, Banks's insights, energy and inherited wealth eventually led to him becoming president of the Royal Society. From his headquarters, as Holmes gracefully phrases it, “his gaze swept steadily round the globe like some vast, enquiring lighthouse beam”. His own days of direct discovery were over, but couldn't other like-minded individuals be encouraged to carry on such wondrous, intense investigations?

One of the young men Banks chose to support was Humphry Davy: friend of the Romantic poets, and — in his quick, intense creation of a safe coal-miner's lamp in response to underground disasters — a man who made himself into a perfect exemplar of the new, Romantic

style of discovery. Davy hurried to the mines, spent intense weeks with the miners and then took himself off to an isolated lab where, using his unique genius, he cracked the problem.

Another of Banks's protégés was William Herschel, the immigrant Hanoverian astronomer. Herschel is most famous now for having measured the orbit of Uranus and establishing this body as a planet, which almost became known as planet George to honour King George III. But Herschel also worked out the shape of the Milky Way and the Sun's off-centre position in it, and he discovered infrared radiation.

Although his work relied on the dull accumulation of observational facts, it was the role of sudden insight and genius that Herschel and others emphasized in their written accounts. In Herschel's case, it was true to his character: he had risen in society by transforming his own life and moving to England. Wouldn't he imagine there were fresh realms — new planets, stars beyond our Solar System, light beyond the visible spectrum — to uncover in nature as well?

Our notion of earlier scientists, including Newton, was rewritten to fit this Romantic view. The young poet William Wordsworth, for example, had famously devalued science

with his harsh line that when we probe the mechanism of a natural process, “we murder to dissect”. But decades later, as the work of Banks and his protégés became better known, Wordsworth shifted to admire science, seeing Newton as the archetypal Romantic hero: “Voyaging through strange seas of Thought, alone.”

Over time, the simplest Romantic imagery slid away. Few of the hundreds of physicists working today at CERN, the European particle-physics lab near Geneva, Switzerland, must view themselves as voyaging anywhere alone. But the

Romantic era left great legacies. Newton's static Universe was gone, and Herschel's dynamic one — in which stars evolved and the heavens were ever-changing — was in its place. This mutable view was useful for the next notable young man, one Charles Darwin, when he was ready to embark on his own voyage of discovery on HMS *Beagle* in 1831. And even among today's doctoral candidates at CERN, who doesn't hope that maybe, with enough concentration, something very special and very unique could still burst out? ■

**David Bodanis** is a writer based in London, and author of *Electric Universe*. His forthcoming book is on the Ten Commandments.



Portrait of Joseph Banks in honour of astronomer William Herschel.

J. RUSSELL/PRIVATE COLLECTION





R. GENDLER/NASA

## Star poetry

**The Very Small Baseline Group Convenes at The Cat and Fiddle**  
by Neil Rollinson

A groaning table of empties makes up our Very Small Array — a barley-scented interferometer. Here we can study the cosmos and drink. We tune in to the microwave sky: to the froth at the edge of the universe. We sup in the dusk, everything glows with its own light: the hedgerow, lawn, the atoms inside the glass. The Milky Way sings in a half-inch of Guinness a song of distant past when the world was a moment old. We gather it all in our mugs, in a pub garden on the edge of the moors, looking down on Jodrell Bank: Queen of the red-light district, cocking her huge lug to the mayhem beyond our patch. The bats are in on it, hunting in ultrasound, catching moths in their fangs, while frogs bark in the meadows, one to the other, a vast unfathomable love-song. I finish my pint and add my glass to the phalanx: the more we drink the clearer we see, as any old soak will tell you. I tip back my head to look at the Pleiades and tumble, arse over tit, into the damp grass. I lie in my cups under the bling of the northern sky. I can hear it now, I can see it all clearly, all and nothing, just the whole sky blazing.

Poets paired with astronomers have created 16 new poems, including this one, that form part of a collection of 100 poems about the wonder of the Universe.

### Dark Matter: Poems of Space

Edited by Maurice Riordan and Jocelyn Bell Burnell

Calouste Gulbenkian Foundation: 2008. 208 pp. £8.50

The pursuit of progress by thinkers and artists became more a wait for a divine spark of inspiration than the steady toil of uncovering that had been accepted by their predecessors. And any means by which that spark could be nurtured was embraced.

To guide readers through the science and culture of this period, Holmes masterfully dips in and out of the life of Joseph Banks. He is an inspired choice. As a curly haired 26-year-old, Banks was aboard HM Bark *Endeavour* on the momentous day in April 1769 when it first glided into sight of Tahiti. Ostensibly Banks was just the expedition's plant collector, but he soon became more fascinated by the island's human inhabitants.

For most of the British crew — young men who had been away from female company for long months — the explorations in Tahiti were of one sort only, with the initial going rate being one ship's nail for one sexual encounter. That rate soon changed — as Holmes describes with gentle skill, the Tahitians were quick to grasp the workings of the market economy, and had a keen eye for the other useful metal objects aboard the ship. Hyperinflation set in, and at one point “there was a crisis when one of the *Endeavour's* crew stole a hundredweight bag of nails, and refused to reveal its whereabouts even after a flogging”.

Banks looked further. He took up Tahitian mistresses too, but systematically recorded the local language, studied their religious systems, and even hinted at the true functional significance of native actions that, at first, seemed to be merely bizarre. Within a few months he had helped set the stage for the modern science of anthropology.

Back in London, Banks's insights, energy and inherited wealth eventually led to him becoming president of the Royal Society. From his headquarters, as Holmes gracefully phrases it, “his gaze swept steadily round the globe like some vast, enquiring lighthouse beam”. His own days of direct discovery were over, but couldn't other like-minded individuals be encouraged to carry on such wondrous, intense investigations?

One of the young men Banks chose to support was Humphry Davy: friend of the Romantic poets, and — in his quick, intense creation of a safe coal-miner's lamp in response to underground disasters — a man who made himself into a perfect exemplar of the new, Romantic

style of discovery. Davy hurried to the mines, spent intense weeks with the miners and then took himself off to an isolated lab where, using his unique genius, he cracked the problem.

Another of Banks's protégés was William Herschel, the immigrant Hanoverian astronomer. Herschel is most famous now for having measured the orbit of Uranus and establishing this body as a planet, which almost became known as planet George to honour King George III. But Herschel also worked out the shape of the Milky Way and the Sun's off-centre position in it, and he discovered infrared radiation.

Although his work relied on the dull accumulation of observational facts, it was the role of sudden insight and genius that Herschel and others emphasized in their written accounts. In Herschel's case, it was true to his character: he had risen in society by transforming his own life and moving to England. Wouldn't he imagine there were fresh realms — new planets, stars beyond our Solar System, light beyond the visible spectrum — to uncover in nature as well?

Our notion of earlier scientists, including Newton, was rewritten to fit this Romantic view. The young poet William Wordsworth, for example, had famously devalued science

with his harsh line that when we probe the mechanism of a natural process, “we murder to dissect”. But decades later, as the work of Banks and his protégés became better known, Wordsworth shifted to admire science, seeing Newton as the archetypal Romantic hero: “Voyaging through strange seas of Thought, alone.”

Over time, the simplest Romantic imagery slid away. Few of the hundreds of physicists working today at CERN, the European particle-physics lab near Geneva, Switzerland, must view themselves as voyaging anywhere alone. But the

Romantic era left great legacies. Newton's static Universe was gone, and Herschel's dynamic one — in which stars evolved and the heavens were ever-changing — was in its place. This mutable view was useful for the next notable young man, one Charles Darwin, when he was ready to embark on his own voyage of discovery on HMS *Beagle* in 1831. And even among today's doctoral candidates at CERN, who doesn't hope that maybe, with enough concentration, something very special and very unique could still burst out?

**David Bodanis** is a writer based in London, and author of *Electric Universe*. His forthcoming book is on the Ten Commandments.



Portrait of Joseph Banks in honour of astronomer William Herschel.

J. RUSSELL/PRIVATE COLLECTION

# Hidden treasures: the Paris Observatory

Giovanni Domenico Cassini helped to create an institution that pinpointed Neptune, showed that light had a finite speed — and even mapped France, explains **Alison Abbott**.

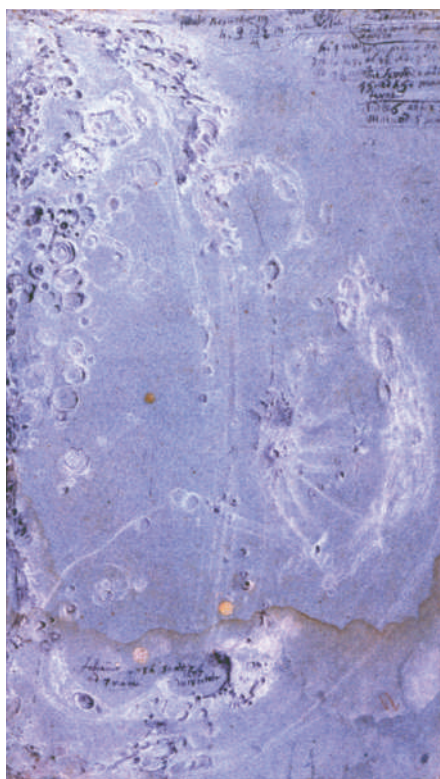


Giovanni Domenico Cassini was a mathematical genius with finely tuned social skills, a quick wit and a habit of enthralling powerful political figures. In 1650, when Cassini was just 25, the Pope appointed him professor of astronomy at his university in Bologna, Italy. He gave Cassini the fattest salary there, to the chagrin of his more established colleagues, and called on him to help resolve some of the highest issues of state. These varied from escorting Sweden's Queen Christina to Rome when she wished to convert to Catholicism — a brief romance apparently sparked between them during the long journey — to determining the direction of water flow in the Apennine mountain range that bordered the papal states, to designing papal fortifications.

Meanwhile, France's Louis XIV, the Sun King, was hatching grandiose plans for his country — among them an academy of sciences and arts and an observatory for experimental sciences, particularly astronomy. He commissioned his minister, Jean-Baptiste Colbert, to scour Europe for the best scientists and artists. In 1668, Colbert made Cassini a financial offer he could not refuse. With the indulgence of the Pope, who was keen to maintain good relations with the French king during a time of political and religious tensions, Cassini set off immediately for Paris. There he joined the fledgling academy and served as royal adviser on the building of the observatory. It was supposed to be a temporary visit. The university in Bologna griped about the Pope's insistence to continue paying him, even when it became clear Cassini wouldn't return. The Pope turned a deaf ear.

Louis XIV loved Cassini from the beginning. Although not formally the first director of the observatory, which was completed in 1682, he assumed the role in practice. Just five years after he arrived, the king bestowed French citizenship on him.

His privileged relationship with Louis XIV was as much a source of annoyance to some Parisian colleagues as his relationship to the Pope had been in Bologna. But Cassini, with his large personality and outstanding scientific achievements — ranging from the discovery of a gap in the rings of Saturn to unveiling the structure of snowflakes — was



Cassini relied on an artist to depict his Moon views.

a major influence on the development of French science.

The king was reportedly less pleased with Cassini's great expedition in the early 1670s. Together with Jean-Felix Picard and Philippe de La Hire, he remapped France using astronomy, ingeniously exploiting Galileo's 1610 discovery of four satellites of Jupiter. They used the eclipses of these satellites as time reference points to measure longitude differences between locations on the coasts of France. The resultant map showed France to be narrower than previously thought, with the kingdom losing 20% of its surface area. Poor recompense, the king said, for the solicitude he had shown his astronomers.

The observatory was built in the countryside, some way from the centre of Paris. Louis XIV had fondly imagined it would be the academy's meeting place, but the scientists found that it took too long to get there and continued to meet in town. Today it has been swallowed up by the city, sheltered from the roar of traffic behind walls that encircle generous

gardens. Visitors will appreciate why Cassini argued, unsuccessfully, with the architects in 1668: the space is hardly optimal for scientific endeavour. It looks more like a palace than a place of work.

But it resounds with history. Vast knowledge was accumulated here, from the 1676 discovery that the speed of light was finite, to the accurate pinpointing of Neptune's position using the observed perturbations of the orbit of Uranus. The high-ceilinged, cavernous rooms are now ideal for displaying the massive wood-and-brass astronomical instruments that form much of the collection of the observatory's museum, created in 1878. Although by then some of the oldest instruments had been lost during expeditions, more than 500 objects remain, including the telescope of Léon Foucault — more famous for his pendulum — and the first to use the thin, silver-plated mirrors found in modern telescopes.

The library, founded in 1785, has also benefited from major acquisitions. During and just after the French Revolution in the 1790s, Cassini's great-grandson donated his forebear's library and extensive records of astronomical observations, as well as those of the three Cassini generations who subsequently directed the observatory — their dynasty lasted for more than a century. Then, the eighteenth-century astronomer Joseph-Nicolas Delisle sold the library a vast number of astronomy manuscripts that he had assiduously collected during his decades of travels around Europe and Russia. Most valuable are the 16 boxes of manuscripts by the great Polish astronomer Johannes Hevelius, including all his correspondence.

The library's greatest treasures are perhaps the original drawings of the surface of the Moon (as pictured) made by Cassini's engraver, who sat beside the astronomer as he made his observations. For all his attributes, Cassini was not a gifted artist. These shaded pastel drawings, made under his guidance and annotated in his own hand, have rarely been seen. Those fortunate enough to examine them will experience a rush — such a direct emotional connection with history is rare, and powerful. ■

**Alison Abbott** is *Nature's* senior European correspondent.

See <http://tinyurl.com/6yxa68> for details. See Editorial, page 7, and [www.nature.com/astro09](http://www.nature.com/astro09).

L'OBSERVATOIRE, PARIS



## NEWS &amp; VIEWS

## QUANTUM PHYSICS

# Squeeze until it hurts

Geoff J. Pryde

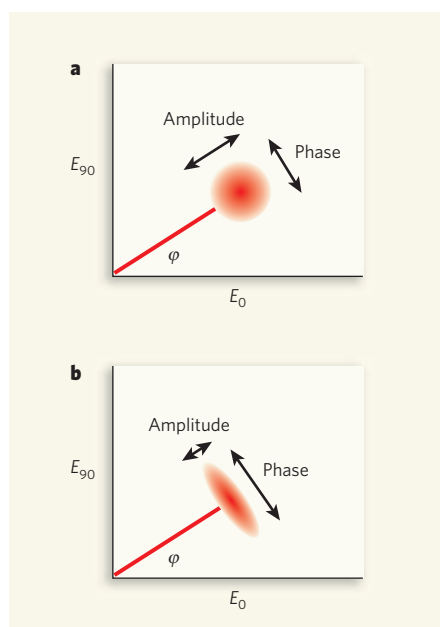
**Quantum systems are uncertain by nature. By ‘squeezing’ this uncertainty, physicists can make better measurements of quantities such as distance. But overdoing it makes things burst out all over the place.**

At the leading edge of experimental science, the latest measurement techniques are promising to provide breakthroughs in our understanding of the Universe. The ever-improving ability to sense small displacements, for example, is at the heart of projects such as the Laser Interferometer Gravitational Wave Observatory (LIGO)<sup>1</sup>, which seeks to observe the faint space-time ripples of distant supernovae. When technical noise is strongly suppressed, the ultimate limit to the precision of any measurement is set by the quantum uncertainty in the measuring system. But even this quantum uncertainty can be reduced — a technique known as ‘squeezing’. On page 67 of this issue, Shalm *et al.*<sup>2</sup> show that squeezing down this quantum uncertainty is not as simple as might be expected — too much squeezing actually worsens measurement precision. Fortunately, they also show that it is possible to recover the best precision allowed by the laws of physics by looking at the ‘over-squeezed’ system in a different way.

That a fundamental limit to measurement precision exists at all is a purely quantum phenomenon. Consider light, the basis of a suite of sensitive interferometric measurement techniques. In the classical picture, light is a wave whose amplitude and phase — where the wave’s peaks and troughs lie — can be specified with infinitesimal precision. But in reality, light has much more character. It is made up of indivisible photons that exhibit probabilistic behaviour when forced to decide which quantum state, out of a range of options presented, to be in.

Shine a hypothetical classical beam on a beam splitter that reflects 50% and transmits 50% of the light, and you know that exactly half goes each way. Put a grainy beam — composed of  $N$  photons, say — on that same beam splitter and it is exactly like tossing a coin  $N$  times. You know that *on average* each outcome should occur  $N/2$  times, but there is some uncertainty because, in a given trial, you might get more heads than tails, for instance.

A picture may help to understand squeezing. Imagine a stick lying in a plane (Fig. 1a) such that its length represents the amplitude of the light and its angle represents the phase. The



**Figure 1 | Quantum uncertainties.** Conceptual representation of the quantum state of light: the length of the red stick represents the amplitude of the light, and its angle  $\phi$  represents the phase. The ‘blob’ represents the quantum uncertainty in knowing the amplitude and phase. The blob is usually a Gaussian probability distribution, and the axes  $E_0$  and  $E_{90}$  represent the quadratures of the electric field  $E$  at frequency  $\omega$ : classically,  $E = E_0 \cos(\omega t) + E_{90} \sin(\omega t)$ . **a**, An unsqueezed state; **b**, a squeezed state. In this example, the amplitude uncertainty is squeezed (decreased) and the phase uncertainty is anti-squeezed (increased).

probabilistic nature of photons smears out the end of the stick, leading to uncertainty in the measurement of phase and amplitude. However, quantum theory tells us we can squeeze this uncertainty ‘blob’ in one direction and it will expand in another direction, just like a balloon filled with water. So, squeezing the uncertainty in amplitude leads to an increased phase uncertainty (Fig. 1b) — but that’s all right if it is just the amplitude that we are trying to measure precisely.

Squeezing itself<sup>3</sup> has been demonstrated many times in a regime in which squeezing harder just makes things better and better.

But it has been predicted that, in systems with limited dimensions, squeezing harder can make things worse<sup>4</sup>. Shalm *et al.*<sup>2</sup> have been able to investigate this fascinating situation experimentally by confining themselves to a quantum system composed of just three photons, which they mash together into the same region of time, space and frequency spectrum.

The quantity that Shalm and colleagues measure is not amplitude or phase, but rather polarization — a property that describes whether the electric-field vector of the light oscillates vertically, horizontally, in a circle, or any combination of the above. These three-photon polarization states are not represented by a stick in a plane, but rather by a stick going from the centre of a sphere to the surface, and the quantum uncertainty is represented by a blob on that surface (Fig. 2a, overleaf). It is now apparent what happens when squeezing occurs: although the uncertainty is reduced in one direction, the blob starts to wrap around the sphere in the other (Fig. 2b). Eventually it goes all the way around and meets up — although the three-fold symmetry leads to a rather more interesting pattern than one might initially expect (Fig. 2c). This over-squeezing is problematic, however. Rather than reducing the uncertainty in the specified direction, the over-squeezing has spread the initial uncertainty into three blobs equally spaced around a ring, meaning that it is now even more uncertain where the end of the stick is — that is to say, what the light’s polarization is.

Does this mean that there is an optimal amount of squeezing — not too little and not too much? If one is limited to detecting all the photons in one bunch (an intensity measurement), the answer is ‘yes’. But when the photons were counted one by one, Shalm *et al.* were able to reveal the correlations between them that give rise to the three fringes around the equator. As this maximally squeezed state is identical to the highly entangled N00N state proposed for quantum metrology<sup>5</sup>, there exist known adaptive-measurement algorithms to extract the complete information in an efficient way<sup>6</sup>.

So where does all of this leave us? Shalm and colleagues have elegantly demonstrated



## 50 YEARS AGO

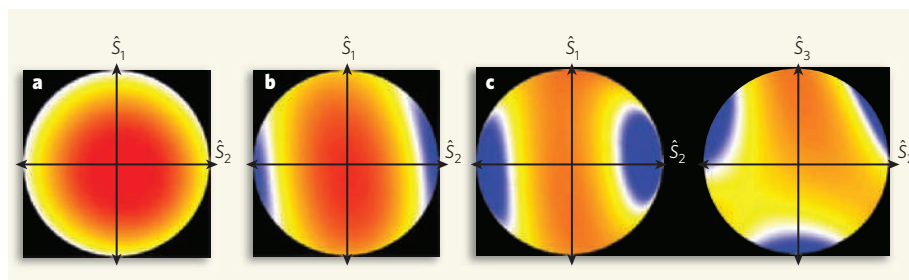
The largest and heaviest satellite so far placed in orbit by the United States was launched in a south-easterly direction from Cape Canaveral in Florida at 23.02 u.t. on December 18. The complete satellite weighs some 8,500 lb., of which only a small proportion, about 150 lb., is pay-load. The remainder of the object is the empty shell of the *Atlas* missile which performed the launching, about 80 ft. long and 10 ft. in diameter. The main purpose of the satellite is to act as a radio relay station, storing and re-transmitting, on frequencies of 132.45 and 139.905 Mc./s., messages sent to it from the ground. It also carries telemetry transmitters operating on 107.97 and 107.94 Mc./s. The satellite is the first to be successfully launched since July 1958, and is designated 1958ζ.

From *Nature* 3 January 1959.

## 100 YEARS AGO

*Fads and Feeding.* By Dr. C. Stanford Read — This is an admirably clear, well reasoned, and sensible little book. One can only hope that it may be widely read and may do something to counteract the ridiculous views on diet which are the result of the cogitation of that dangerous class of people who, having a little knowledge, supply the remainder from their own imagination. In spite of the importance of a suitable diet for health, there is perhaps no other subject which breeds so many fads ... The key-word of Dr. Read's book is moderation; moderation in meat-eating, in tea-drinking, in the use of alcoholic drinks and the like. He is also moderate in the way he deals with the faddists, the vegetarians, the uric-acid-free dietists, the teetotalers, and the rest. Perhaps this method of dealing with them is the most effectual with the public, who, taken as a whole, are moderate, and temperate too.

From *Nature* 31 December 1908.



**Figure 2 | Squeezing a triphoton.** The position of the quantum state on the surface of a sphere — seen here in two-dimensional projections — describes the polarization of a three-photon (triphoton) state. The directions  $\hat{S}_1$ ,  $\hat{S}_2$  and  $\hat{S}_3$  represent the degree of horizontal, 45° and right-circular polarizations, respectively. The colours correspond to different levels of probability for the polarization state (red for highest probability, blue for lowest); the red and yellow regions represent the quantum uncertainty blob. The uncertainty in the  $\hat{S}_2$  direction is reduced as the squeezing is increased (a–c). **a**, An unsqueezed state. **b**, As the squeezing is increased, the uncertainty in the  $\hat{S}_2$  direction is decreased at the expense of increased uncertainty in the  $\hat{S}_1$  direction. **c**, Shalm *et al.*<sup>2</sup> show that in the over-squeezed case the uncertainty blob completely wraps around the sphere. Although the regions of uncertainty are small in the  $\hat{S}_2$  direction, there are three of them — as shown in an alternative projection (right) — leading to a large overall uncertainty. (Modified from ref. 2.)

the connection between highly entangled states and the squeezing of quantum states by showing the continuum of quantum states with reduced measurement uncertainty. And this idea does not only apply to photons. Recently, similar effects were observed in an atomic spin system<sup>7</sup>. So, are physicists now the masters of quantum uncertainty? Well, not quite. It remains a difficult proposition to highly squeeze large numbers of quantum systems, and a few photons is a long way from the large entangled states required for practical application of quantum-enhanced precision measurement. But our control of the quantum world is always improving, and we may one day see optimum-precision measurements with large ensembles. In the meantime, we can look for applications of those squeezed

and entangled states that can be made<sup>8–10</sup>. And we can admire the peculiar beauty and symmetries of the quantum world. ■

Geoff J. Pryde is at the Centre for Quantum Dynamics, Griffith University, Brisbane, Queensland 4111, Australia.  
e-mail: g.pryde@griffith.edu.au

1. [www.ligo.caltech.edu](http://www.ligo.caltech.edu)
2. Shalm, L. K., Adamson, R. B. A. & Steinberg, A. M. *Nature* **457**, 67–70 (2009).
3. Walls, D. F. & Milburn, G. J. *Quantum Optics* (Springer, 1994).
4. Combes, J. & Wiseman, H. M. *J. Opt. B: Quant. Semiclass. Opt.* **7**, 14–21 (2005).
5. Bouwmeester, D. *Nature* **429**, 139–141 (2004).
6. Higgins, B. L., Berry, D. W., Bartlett, S. D., Wiseman, H. M. & Pryde, G. J. *Nature* **450**, 393–396 (2007).
7. Chaudhury, S. *et al. Phys. Rev. Lett.* **99**, 163002 (2007).
8. McKenzie, K. *et al. Phys. Rev. Lett.* **93**, 161105 (2004).
9. Goda, K. *et al. Nature Phys.* **4**, 472–476 (2008).
10. Meyer, V. *et al. Phys. Rev. Lett.* **86**, 5870–5873 (2001).

## CANCER

# Inflaming metastasis

Alberto Mantovani

**Cancer can be defined by six hallmarks, including uncontrollable growth, immortality and the ability to invade other tissues. Increasing evidence suggests that a seventh feature should make this list — inflammation.**

Malignant tumours are characterized by their ability to metastasize, that is, to invade anatomically distant normal tissues and to seed and grow there. During this complex and highly selective process, tumour cells leave their primary site and disseminate by various routes, such as the blood and lymph vessels. Not all cancer cells can metastasize<sup>1</sup>, because successful metastasis depends both on intrinsic properties of the tumour cells and on factors derived from the tumour microenvironment. For example, the microenvironment provides blood and lymphatic vessels in and around the

tumour, an inflammatory milieu consisting of immune cells and their secretory products, and a scaffold in the form of the extracellular matrix for further growth. Writing in this issue, Kim *et al.*<sup>2</sup> shed light on the unexpected molecular pathways that link inflammation in the tumour microenvironment to metastasis (page 102).

The link between inflammation and cancer is well documented<sup>3,4</sup>. Several inflammatory diseases, including inflammatory bowel disease, increase the risk of cancer. Conversely, in tumours that are epidemiologically unrelated





## 50 YEARS AGO

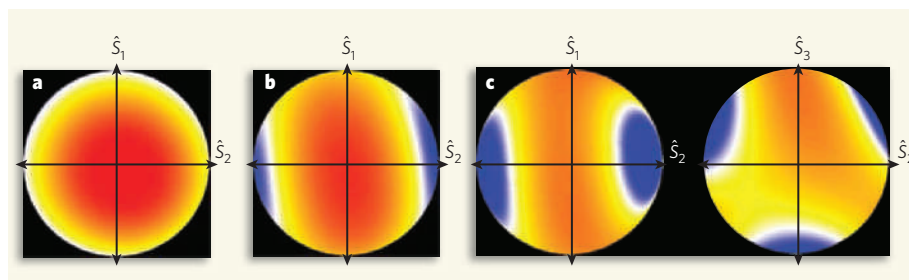
The largest and heaviest satellite so far placed in orbit by the United States was launched in a south-easterly direction from Cape Canaveral in Florida at 23.02 u.t. on December 18. The complete satellite weighs some 8,500 lb., of which only a small proportion, about 150 lb., is pay-load. The remainder of the object is the empty shell of the *Atlas* missile which performed the launching, about 80 ft. long and 10 ft. in diameter. The main purpose of the satellite is to act as a radio relay station, storing and re-transmitting, on frequencies of 132.45 and 139.905 Mc./s., messages sent to it from the ground. It also carries telemetry transmitters operating on 107.97 and 107.94 Mc./s. The satellite is the first to be successfully launched since July 1958, and is designated 1958ζ.

From *Nature* 3 January 1959.

## 100 YEARS AGO

*Fads and Feeding.* By Dr. C. Stanford Read — This is an admirably clear, well reasoned, and sensible little book. One can only hope that it may be widely read and may do something to counteract the ridiculous views on diet which are the result of the cogitation of that dangerous class of people who, having a little knowledge, supply the remainder from their own imagination. In spite of the importance of a suitable diet for health, there is perhaps no other subject which breeds so many fads ... The key-word of Dr. Read's book is moderation; moderation in meat-eating, in tea-drinking, in the use of alcoholic drinks and the like. He is also moderate in the way he deals with the faddists, the vegetarians, the uric-acid-free dietists, the teetotalers, and the rest. Perhaps this method of dealing with them is the most effectual with the public, who, taken as a whole, are moderate, and temperate too.

From *Nature* 31 December 1908.



**Figure 2 | Squeezing a triphoton.** The position of the quantum state on the surface of a sphere — seen here in two-dimensional projections — describes the polarization of a three-photon (triphoton) state. The directions  $\hat{S}_1$ ,  $\hat{S}_2$  and  $\hat{S}_3$  represent the degree of horizontal, 45° and right-circular polarizations, respectively. The colours correspond to different levels of probability for the polarization state (red for highest probability, blue for lowest); the red and yellow regions represent the quantum uncertainty blob. The uncertainty in the  $\hat{S}_2$  direction is reduced as the squeezing is increased (a–c). **a**, An unsqueezed state. **b**, As the squeezing is increased, the uncertainty in the  $\hat{S}_2$  direction is decreased at the expense of increased uncertainty in the  $\hat{S}_1$  direction. **c**, Shalm *et al.*<sup>2</sup> show that in the over-squeezed case the uncertainty blob completely wraps around the sphere. Although the regions of uncertainty are small in the  $\hat{S}_2$  direction, there are three of them — as shown in an alternative projection (right) — leading to a large overall uncertainty. (Modified from ref. 2.)

the connection between highly entangled states and the squeezing of quantum states by showing the continuum of quantum states with reduced measurement uncertainty. And this idea does not only apply to photons. Recently, similar effects were observed in an atomic spin system<sup>7</sup>. So, are physicists now the masters of quantum uncertainty? Well, not quite. It remains a difficult proposition to highly squeeze large numbers of quantum systems, and a few photons is a long way from the large entangled states required for practical application of quantum-enhanced precision measurement. But our control of the quantum world is always improving, and we may one day see optimum-precision measurements with large ensembles. In the meantime, we can look for applications of those squeezed

and entangled states that can be made<sup>8–10</sup>. And we can admire the peculiar beauty and symmetries of the quantum world. ■

Geoff J. Pryde is at the Centre for Quantum Dynamics, Griffith University, Brisbane, Queensland 4111, Australia.  
e-mail: g.pryde@griffith.edu.au

1. [www.ligo.caltech.edu](http://www.ligo.caltech.edu)
2. Shalm, L. K., Adamson, R. B. A. & Steinberg, A. M. *Nature* **457**, 67–70 (2009).
3. Walls, D. F. & Milburn, G. J. *Quantum Optics* (Springer, 1994).
4. Combes, J. & Wiseman, H. M. *J. Opt. B: Quant. Semiclass. Opt.* **7**, 14–21 (2005).
5. Bouwmeester, D. *Nature* **429**, 139–141 (2004).
6. Higgins, B. L., Berry, D. W., Bartlett, S. D., Wiseman, H. M. & Pryde, G. J. *Nature* **450**, 393–396 (2007).
7. Chaudhury, S. *et al. Phys. Rev. Lett.* **99**, 163002 (2007).
8. McKenzie, K. *et al. Phys. Rev. Lett.* **93**, 161105 (2004).
9. Goda, K. *et al. Nature Phys.* **4**, 472–476 (2008).
10. Meyer, V. *et al. Phys. Rev. Lett.* **86**, 5870–5873 (2001).

## CANCER

# Inflaming metastasis

Alberto Mantovani

**Cancer can be defined by six hallmarks, including uncontrollable growth, immortality and the ability to invade other tissues. Increasing evidence suggests that a seventh feature should make this list — inflammation.**

Malignant tumours are characterized by their ability to metastasize, that is, to invade anatomically distant normal tissues and to seed and grow there. During this complex and highly selective process, tumour cells leave their primary site and disseminate by various routes, such as the blood and lymph vessels. Not all cancer cells can metastasize<sup>1</sup>, because successful metastasis depends both on intrinsic properties of the tumour cells and on factors derived from the tumour microenvironment. For example, the microenvironment provides blood and lymphatic vessels in and around the

tumour, an inflammatory milieu consisting of immune cells and their secretory products, and a scaffold in the form of the extracellular matrix for further growth. Writing in this issue, Kim *et al.*<sup>2</sup> shed light on the unexpected molecular pathways that link inflammation in the tumour microenvironment to metastasis (page 102).

The link between inflammation and cancer is well documented<sup>3,4</sup>. Several inflammatory diseases, including inflammatory bowel disease, increase the risk of cancer. Conversely, in tumours that are epidemiologically unrelated



## 50 YEARS AGO

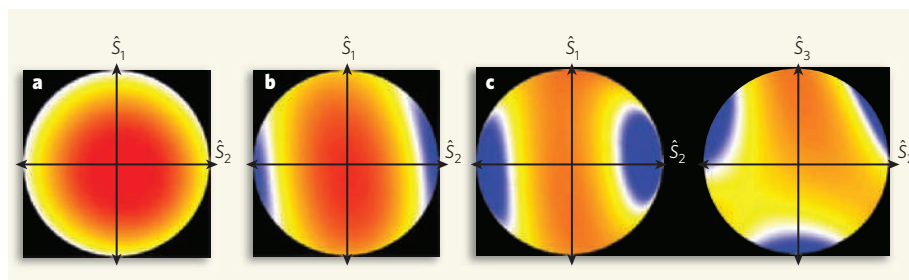
The largest and heaviest satellite so far placed in orbit by the United States was launched in a south-easterly direction from Cape Canaveral in Florida at 23.02 u.t. on December 18. The complete satellite weighs some 8,500 lb., of which only a small proportion, about 150 lb., is pay-load. The remainder of the object is the empty shell of the *Atlas* missile which performed the launching, about 80 ft. long and 10 ft. in diameter. The main purpose of the satellite is to act as a radio relay station, storing and re-transmitting, on frequencies of 132.45 and 139.905 Mc./s., messages sent to it from the ground. It also carries telemetry transmitters operating on 107.97 and 107.94 Mc./s. The satellite is the first to be successfully launched since July 1958, and is designated 1958ζ.

From *Nature* 3 January 1959.

## 100 YEARS AGO

*Fads and Feeding.* By Dr. C. Stanford Read — This is an admirably clear, well reasoned, and sensible little book. One can only hope that it may be widely read and may do something to counteract the ridiculous views on diet which are the result of the cogitation of that dangerous class of people who, having a little knowledge, supply the remainder from their own imagination. In spite of the importance of a suitable diet for health, there is perhaps no other subject which breeds so many fads ... The key-word of Dr. Read's book is moderation; moderation in meat-eating, in tea-drinking, in the use of alcoholic drinks and the like. He is also moderate in the way he deals with the faddists, the vegetarians, the uric-acid-free dietists, the teetotalers, and the rest. Perhaps this method of dealing with them is the most effectual with the public, who, taken as a whole, are moderate, and temperate too.

From *Nature* 31 December 1908.



**Figure 2 | Squeezing a triphoton.** The position of the quantum state on the surface of a sphere — seen here in two-dimensional projections — describes the polarization of a three-photon (triphoton) state. The directions  $\hat{S}_1$ ,  $\hat{S}_2$  and  $\hat{S}_3$  represent the degree of horizontal, 45° and right-circular polarizations, respectively. The colours correspond to different levels of probability for the polarization state (red for highest probability, blue for lowest); the red and yellow regions represent the quantum uncertainty blob. The uncertainty in the  $\hat{S}_2$  direction is reduced as the squeezing is increased (a–c). **a**, An unsqueezed state. **b**, As the squeezing is increased, the uncertainty in the  $\hat{S}_2$  direction is decreased at the expense of increased uncertainty in the  $\hat{S}_1$  direction. **c**, Shalm *et al.*<sup>2</sup> show that in the over-squeezed case the uncertainty blob completely wraps around the sphere. Although the regions of uncertainty are small in the  $\hat{S}_2$  direction, there are three of them — as shown in an alternative projection (right) — leading to a large overall uncertainty. (Modified from ref. 2.)

the connection between highly entangled states and the squeezing of quantum states by showing the continuum of quantum states with reduced measurement uncertainty. And this idea does not only apply to photons. Recently, similar effects were observed in an atomic spin system<sup>7</sup>. So, are physicists now the masters of quantum uncertainty? Well, not quite. It remains a difficult proposition to highly squeeze large numbers of quantum systems, and a few photons is a long way from the large entangled states required for practical application of quantum-enhanced precision measurement. But our control of the quantum world is always improving, and we may one day see optimum-precision measurements with large ensembles. In the meantime, we can look for applications of those squeezed

and entangled states that can be made<sup>8–10</sup>. And we can admire the peculiar beauty and symmetries of the quantum world. ■

Geoff J. Pryde is at the Centre for Quantum Dynamics, Griffith University, Brisbane, Queensland 4111, Australia.  
e-mail: g.pryde@griffith.edu.au

1. www.ligo.caltech.edu
2. Shalm, L. K., Adamson, R. B. A. & Steinberg, A. M. *Nature* **457**, 67–70 (2009).
3. Walls, D. F. & Milburn, G. J. *Quantum Optics* (Springer, 1994).
4. Combes, J. & Wiseman, H. M. *J. Opt. B: Quant. Semiclass. Opt.* **7**, 14–21 (2005).
5. Bouwmeester, D. *Nature* **429**, 139–141 (2004).
6. Higgins, B. L., Berry, D. W., Bartlett, S. D., Wiseman, H. M. & Pryde, G. J. *Nature* **450**, 393–396 (2007).
7. Chaudhury, S. *et al.* *Phys. Rev. Lett.* **99**, 163002 (2007).
8. McKenzie, K. *et al.* *Phys. Rev. Lett.* **93**, 161105 (2004).
9. Goda, K. *et al.* *Nature Phys.* **4**, 472–476 (2008).
10. Meyer, V. *et al.* *Phys. Rev. Lett.* **86**, 5870–5873 (2001).

## CANCER

# Inflaming metastasis

Alberto Mantovani

**Cancer can be defined by six hallmarks, including uncontrollable growth, immortality and the ability to invade other tissues. Increasing evidence suggests that a seventh feature should make this list — inflammation.**

Malignant tumours are characterized by their ability to metastasize, that is, to invade anatomically distant normal tissues and to seed and grow there. During this complex and highly selective process, tumour cells leave their primary site and disseminate by various routes, such as the blood and lymph vessels. Not all cancer cells can metastasize<sup>1</sup>, because successful metastasis depends both on intrinsic properties of the tumour cells and on factors derived from the tumour microenvironment. For example, the microenvironment provides blood and lymphatic vessels in and around the

tumour, an inflammatory milieu consisting of immune cells and their secretory products, and a scaffold in the form of the extracellular matrix for further growth. Writing in this issue, Kim *et al.*<sup>2</sup> shed light on the unexpected molecular pathways that link inflammation in the tumour microenvironment to metastasis (page 102).

The link between inflammation and cancer is well documented<sup>3,4</sup>. Several inflammatory diseases, including inflammatory bowel disease, increase the risk of cancer. Conversely, in tumours that are epidemiologically unrelated

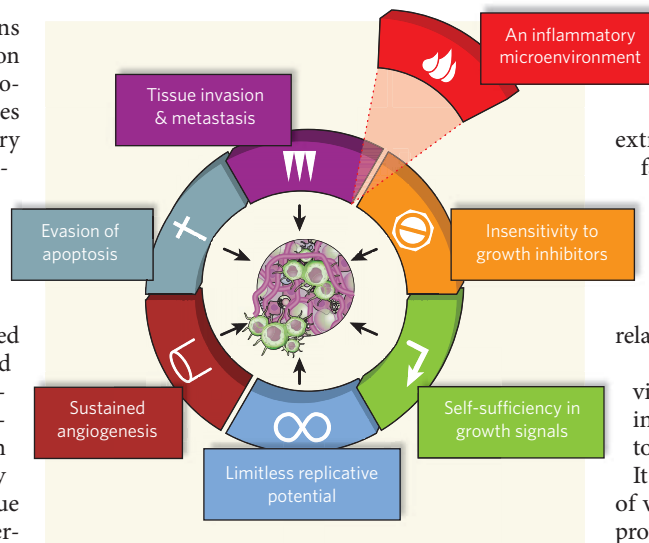


to overt inflammatory conditions (such as breast cancer), the activation of oncogenes can orchestrate the production of inflammatory molecules and the recruitment of inflammatory cells. In the tumour microenvironment, inflammatory cells and molecules influence almost every aspect of cancer progress, including the tumour cells' ability to metastasize<sup>3</sup>. Thus, whereas there were previously six recognized hallmarks of cancer — unlimited replicative potential, self-sufficiency in growth signals, insensitivity to growth inhibitors, evasion of programmed cell death, ability to develop blood vessels, and tissue invasion and metastasis<sup>5</sup> — cancer-related inflammation now emerges as number seven (Fig. 1).

A group of cytokine proteins, including IL-1, IL-6, TNF and RANKL, activate inflammation and are known to augment tumour cells' ability to metastasize by affecting several steps in the cells' dissemination and implantation at secondary sites<sup>3,6,7</sup>. Inflammatory cytokines lie downstream of the 'master' gene transcription factor for promoting inflammation — NF- $\kappa$ B — which is itself activated by them<sup>3</sup>. A major source of inflammatory cytokines in the tumour microenvironment are specialized white blood cells called macrophages. Tumour-associated macrophages assist the malignant behaviour of tumour cells, not just by producing cytokines, but also by secreting growth factors and matrix-degrading enzymes<sup>8–10</sup>.

Kim *et al.*<sup>2</sup> explored the molecular pathways linking tumour cells, macrophages and metastasis. By purifying the components of the medium in which the tumour cells (the Lewis lung carcinoma cell line) were grown, they isolated a factor that induced cytokine production by macrophages. They identified this tumour-derived macrophage activator as versican, a protein of the extracellular matrix that is frequently upregulated in human tumours. The authors found that versican is recognized by TLR2 and TLR6, two receptor proteins that belong to a family of cellular sensors of microbially derived molecules and tissue damage. They then went on to silence versican in tumour cells by an RNA interference technique, and to use mice in which TNF and TLR were absent. On the basis of the evidence obtained, the authors propose that, in the Lewis lung carcinoma model, tumour-derived versican acts on macrophages through TLR2/TLR6, leading to the production of inflammatory cytokines, which enhance metastasis.

Kim and colleagues' observations highlight the importance of the extracellular matrix in cancer-related inflammation. The matrix acts as a depot of cytokines and growth factors, in



**Figure 1 | The hallmarks of cancer.** In 2000, Hanahan and Weinberg<sup>5</sup> proposed a model to define the six properties that a tumour acquires. These are unlimited replicative potential, ability to develop blood vessels (angiogenesis), evasion of programmed cell death (apoptosis), self-sufficiency in growth signals, insensitivity to inhibitors of growth, and tissue invasion and metastasis. Kim and colleagues' findings<sup>2</sup>, together with those of other studies<sup>3,4</sup>, indicate that this model should be revised to include cancer-related inflammation as an additional hallmark. (Adapted from ref. 5.)

particular vascular endothelial growth factor, which is mobilized by enzymes originating from inflammatory white blood cells and which promotes blood-vessel formation during tumour progression<sup>4,5</sup>. Moreover, during the development of cancers caused by human papillomavirus, immune cells called B cells

orchestrate inflammation remotely by producing antibodies that become localized in the extracellular matrix<sup>11</sup>.

What's more, a macrophage-derived extracellular-matrix protein called SPARC facilitates tumour-cell motility and metastasis<sup>12</sup>. So it seems that extracellular-matrix components are much more than a scaffold, or a substrate to be consumed during tumour-cell invasion, but instead represent a central component of cancer-related inflammation.

The present study<sup>2</sup> offers unexpected vistas on the molecular pathways that link inflammation to acquisition of the capacity to metastasize during tumour progression.

It will be essential to assess the significance of versican and other extracellular-matrix proteins in models that reflect the diversity of human cancer, for from such work innovative therapeutic strategies may follow.

Alberto Mantovani is at the Istituto Clinico Humanitas IRCCS, and the University of Milan, Rozzano, Milan 20089, Italy.

e-mail: alberto.mantovani@humanitas.it

1. Fidler, I. J. *Nature Rev. Cancer* **3**, 453–458 (2003).
2. Kim, S. *et al. Nature* **457**, 102–106 (2009).
3. Mantovani, A., Allavena, P., Sica, A. & Balkwill, F. *Nature* **454**, 436–444 (2008).
4. Coussens, L. M. & Werb, Z. *Nature* **420**, 860–867 (2002).
5. Hanahan, D. & Weinberg, R. A. *Cell* **100**, 57–70 (2000).
6. Giavazzi, R. *et al. Cancer Res.* **50**, 4771–4775 (1990).
7. Luo, J. L. *et al. Nature* **446**, 690–694 (2007).
8. Wyckoff, J. B. *et al. Cancer Res.* **67**, 2649–2656 (2007).
9. Mantovani, A., Schioppa, T., Porta, C., Allavena, P. & Sica, A. *Cancer Metastasis Rev.* **25**, 315–322 (2006).
10. Yang, L. *et al. Cancer Cell* **13**, 23–35 (2008).
11. de Visser, K. E., Korets, L. V. & Coussens, L. M. *Cancer Cell* **7**, 411–423 (2005).
12. Sangaletti, S. *et al. Cancer Res.* **68**, 9050–9059 (2008).

## ASTROPHYSICS

# Star formation branches out

Ralph E. Pudritz

**Deciphering how stars form within turbulent, dense clouds of molecular gas has been a challenge. An innovative technique that uses a tree diagram provides insight into the process.**

An understanding of how stars form has been hampered by the complexity of the clouds of cold molecular gas within which their formation occurs. To elucidate this process, the effects of gravity must be traced across a wide range of scales, particularly at the large distances over which it operates in these gas clouds. On page 63 of this issue, Goodman *et al.*<sup>1</sup> show how a hierarchical tree diagram — a 'dendrogram' — can be used to disentangle the gravitational connections that tie the gas together on many different scales.

The movements of gas in molecular clouds are measured by spectral shifts. Depending on whether a parcel of gas is moving towards or

away from the observer, the wavelength of the millimetre emission from a molecule such as carbon monoxide is shifted towards shorter ('blueshifted') or longer ('redshifted') wavelengths. By measuring these Doppler shifts at each point on the sky, one can determine the relative velocities of parcels of gas in the cloud along each line of sight. It turns out that the gas motions in such clouds are mainly highly supersonic. Indeed, computer simulations show<sup>2</sup> that the network of dense filamentary structures seen in clouds is probably a direct consequence of such supersonic gas flows (Fig. 1).

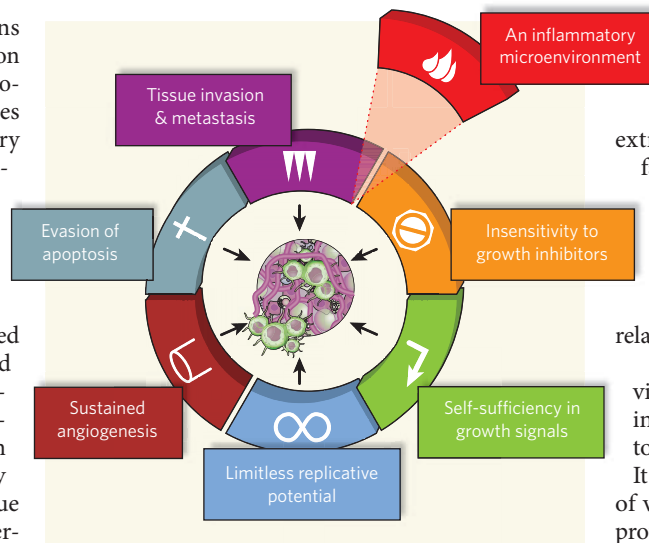
As with many astronomical observations,

to overt inflammatory conditions (such as breast cancer), the activation of oncogenes can orchestrate the production of inflammatory molecules and the recruitment of inflammatory cells. In the tumour microenvironment, inflammatory cells and molecules influence almost every aspect of cancer progress, including the tumour cells' ability to metastasize<sup>3</sup>. Thus, whereas there were previously six recognized hallmarks of cancer — unlimited replicative potential, self-sufficiency in growth signals, insensitivity to growth inhibitors, evasion of programmed cell death, ability to develop blood vessels, and tissue invasion and metastasis<sup>5</sup> — cancer-related inflammation now emerges as number seven (Fig. 1).

A group of cytokine proteins, including IL-1, IL-6, TNF and RANKL, activate inflammation and are known to augment tumour cells' ability to metastasize by affecting several steps in the cells' dissemination and implantation at secondary sites<sup>3,6,7</sup>. Inflammatory cytokines lie downstream of the 'master' gene transcription factor for promoting inflammation — NF- $\kappa$ B — which is itself activated by them<sup>3</sup>. A major source of inflammatory cytokines in the tumour microenvironment are specialized white blood cells called macrophages. Tumour-associated macrophages assist the malignant behaviour of tumour cells, not just by producing cytokines, but also by secreting growth factors and matrix-degrading enzymes<sup>8–10</sup>.

Kim *et al.*<sup>2</sup> explored the molecular pathways linking tumour cells, macrophages and metastasis. By purifying the components of the medium in which the tumour cells (the Lewis lung carcinoma cell line) were grown, they isolated a factor that induced cytokine production by macrophages. They identified this tumour-derived macrophage activator as versican, a protein of the extracellular matrix that is frequently upregulated in human tumours. The authors found that versican is recognized by TLR2 and TLR6, two receptor proteins that belong to a family of cellular sensors of microbially derived molecules and tissue damage. They then went on to silence versican in tumour cells by an RNA interference technique, and to use mice in which TNF and TLR were absent. On the basis of the evidence obtained, the authors propose that, in the Lewis lung carcinoma model, tumour-derived versican acts on macrophages through TLR2/TLR6, leading to the production of inflammatory cytokines, which enhance metastasis.

Kim and colleagues' observations highlight the importance of the extracellular matrix in cancer-related inflammation. The matrix acts as a depot of cytokines and growth factors, in



**Figure 1 | The hallmarks of cancer.** In 2000, Hanahan and Weinberg<sup>5</sup> proposed a model to define the six properties that a tumour acquires. These are unlimited replicative potential, ability to develop blood vessels (angiogenesis), evasion of programmed cell death (apoptosis), self-sufficiency in growth signals, insensitivity to inhibitors of growth, and tissue invasion and metastasis. Kim and colleagues' findings<sup>2</sup>, together with those of other studies<sup>3,4</sup>, indicate that this model should be revised to include cancer-related inflammation as an additional hallmark. (Adapted from ref. 5.)

particular vascular endothelial growth factor, which is mobilized by enzymes originating from inflammatory white blood cells and which promotes blood-vessel formation during tumour progression<sup>4,5</sup>. Moreover, during the development of cancers caused by human papillomavirus, immune cells called B cells

orchestrate inflammation remotely by producing antibodies that become localized in the extracellular matrix<sup>11</sup>.

What's more, a macrophage-derived extracellular-matrix protein called SPARC facilitates tumour-cell motility and metastasis<sup>12</sup>. So it seems that extracellular-matrix components are much more than a scaffold, or a substrate to be consumed during tumour-cell invasion, but instead represent a central component of cancer-related inflammation.

The present study<sup>2</sup> offers unexpected vistas on the molecular pathways that link inflammation to acquisition of the capacity to metastasize during tumour progression.

It will be essential to assess the significance of versican and other extracellular-matrix proteins in models that reflect the diversity of human cancer, for from such work innovative therapeutic strategies may follow.

Alberto Mantovani is at the Istituto Clinico Humanitas IRCCS, and the University of Milan, Rozzano, Milan 20089, Italy.

e-mail: alberto.mantovani@humanitas.it

1. Fidler, I. J. *Nature Rev. Cancer* **3**, 453–458 (2003).
2. Kim, S. *et al. Nature* **457**, 102–106 (2009).
3. Mantovani, A., Allavena, P., Sica, A. & Balkwill, F. *Nature* **454**, 436–444 (2008).
4. Coussens, L. M. & Werb, Z. *Nature* **420**, 860–867 (2002).
5. Hanahan, D. & Weinberg, R. A. *Cell* **100**, 57–70 (2000).
6. Giavazzi, R. *et al. Cancer Res.* **50**, 4771–4775 (1990).
7. Luo, J. L. *et al. Nature* **446**, 690–694 (2007).
8. Wyckoff, J. B. *et al. Cancer Res.* **67**, 2649–2656 (2007).
9. Mantovani, A., Schioppa, T., Porta, C., Allavena, P. & Sica, A. *Cancer Metastasis Rev.* **25**, 315–322 (2006).
10. Yang, L. *et al. Cancer Cell* **13**, 23–35 (2008).
11. de Visser, K. E., Korets, L. V. & Coussens, L. M. *Cancer Cell* **7**, 411–423 (2005).
12. Sangaletti, S. *et al. Cancer Res.* **68**, 9050–9059 (2008).

## ASTROPHYSICS

# Star formation branches out

Ralph E. Pudritz

**Deciphering how stars form within turbulent, dense clouds of molecular gas has been a challenge. An innovative technique that uses a tree diagram provides insight into the process.**

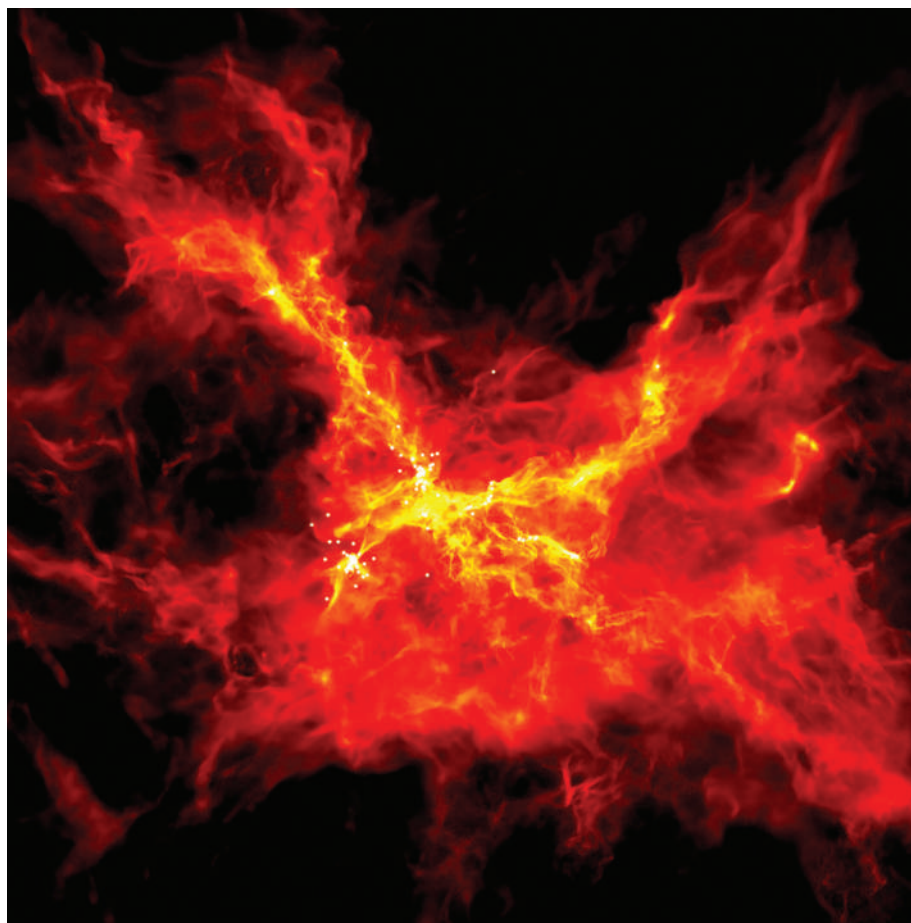
An understanding of how stars form has been hampered by the complexity of the clouds of cold molecular gas within which their formation occurs. To elucidate this process, the effects of gravity must be traced across a wide range of scales, particularly at the large distances over which it operates in these gas clouds. On page 63 of this issue, Goodman *et al.*<sup>1</sup> show how a hierarchical tree diagram — a 'dendrogram' — can be used to disentangle the gravitational connections that tie the gas together on many different scales.

The movements of gas in molecular clouds are measured by spectral shifts. Depending on whether a parcel of gas is moving towards or

away from the observer, the wavelength of the millimetre emission from a molecule such as carbon monoxide is shifted towards shorter ('blueshifted') or longer ('redshifted') wavelengths. By measuring these Doppler shifts at each point on the sky, one can determine the relative velocities of parcels of gas in the cloud along each line of sight. It turns out that the gas motions in such clouds are mainly highly supersonic. Indeed, computer simulations show<sup>2</sup> that the network of dense filamentary structures seen in clouds is probably a direct consequence of such supersonic gas flows (Fig. 1).

As with many astronomical observations,





**Figure 1 | Stellar nursery.** The image shows a computer simulation<sup>2</sup> of the formation of stars within a turbulent, self-gravitating cloud of gas. The initial mass of this star-cluster-forming cloud, which is modelled as a sphere of uniform density, is 500 solar masses. The sphere's radius is 83,300 astronomical units (1 AU is the mean distance between Earth and the Sun) and its temperature is 10 kelvin. Supersonic turbulence compresses the gas into many filaments and smaller, dense regions. The simulation is viewed after one free-fall time — the time taken for a gas parcel to collapse freely to the cloud centre — has elapsed, which is  $1.9 \times 10^5$  years for these simulation parameters. The white dots correspond to small, dense gas 'cores' that collapse to form individual stars. These would correspond physically to regions such as those denoted by the billiard balls in Figure 1 of the paper by Goodman *et al.*<sup>1</sup> (page 63).

however, we do not know the distance of any object (gas in this case) from Earth without using further painstaking methods. In observing molecular clouds, one is limited to measuring the two-dimensional position of the total gas emission on the sky. This measurement involves two coordinates that are akin to latitude and longitude on Earth's surface, as well as the relative velocity of gas at that position. Thus, a map of such a cloud is a sequence of position–position–velocity ( $p$ – $p$ – $v$ ) measurements of the gas across the whole cloud. But without the ability to observe the full, three-dimensional gas cloud, how can its true structure be deduced, let alone the strength of the various forces that control where and when stars will form?

The approach usually taken to tackle this problem consists of segmenting the clumpy cloud into a collection of structures (clumps) using a computer program called CLUMPFIND. The end result is analogous to a topographical relief map of a mountain range. Such a map typically shows peaks that stand

out from ridges or are isolated, and provide a series of contours that demarcate different elevations. If one now decided to break the range up into discrete 'mountains', one would identify the peaks and, using the various contour levels, try to decide whether smaller outcrops 'belonged' to a given mountain or were independent structures. In a  $p$ – $p$ – $v$  map of a molecular cloud, it is the contours and peaks of gas-column density — the sum of the emission from all gas parcels moving at a given velocity along a given line of sight — that play the role of topological relief in this mountain-range analogy.

The problem with this approach arises as soon as the results are used to try to provide insight into how stars form. For example, the column densities allow one to measure the masses of the clumps. One can then count the number of clumps with a given mass. The resulting distribution of clump masses is used to work out how star formation might occur<sup>3,4</sup>. With programs similar to CLUMPFIND, the data suggest that the

clump-mass distribution closely resembles the distribution of star masses. A plausible but debated inference is that the origin of stellar masses may derive directly from the turbulent process that produced the clumps.

The fly in this ointment, however, as Goodman *et al.*<sup>1</sup> show (see Fig. 1 of their Supplementary Information), is that if one adopts different threshold levels for contouring the maps, the column-density distribution changes. This is similar to the situation in a topographical relief map of a mountain range: the list of discrete mountains and their properties depends on the choice of threshold level picked to define mountains and smaller outcrops. This is unsettling — such an approach does not provide a completely objective way of measuring the actual distribution of column densities.

Enter the dendrogram technique advocated by the authors<sup>1</sup>. Rather than dividing the  $p$ – $p$ – $v$  data into a priori distinct structures in a subjective way, they use a method that is sensitive to the structures' intrinsic hierarchy (structure within structure). The data are broken up into 'leaves', 'branches' and 'trunks'. Leaves are identified as sufficiently strong maxima in the column-density maps, and connections between them are made by branches (their environs). The collection of physically related branches defines a trunk.

Every point on the dendrogram corresponds to a closed 'isosurface' on the map that encloses one or more column-density maxima (see Fig. 2 on page 64). The authors then show how physical properties can be ascribed to regions within these isosurfaces. A critical property is the 'virial' parameter — the ratio of the energy of the gas motions (which depends on the gas velocity in the region) to the gravitational energy (which depends on the mass and size of the region). If this ratio is sufficiently small, the gas in that region will be self-gravitating and prone to form stars. The dendrogram thus traces the relative strength of the gravitational force across the gas cloud.

Application of the dendrogram technique to observations of the gas cloud L1448, and to computer simulations in which only turbulence — and not self-gravity — is taken into account, shows inconsistencies. In contrast to the simulations, in which most of the gas is found to be self-gravitating on all spatial scales, the observations show that, although a large fraction of the gas is self-gravitating at large scales, at smaller scales that fraction is much lower.

Interestingly, strong local column-density maxima — which correspond to the dense gas 'cores' in which stars are observed to form — are sparser in Goodman and colleagues' dendrogram than those found with the CLUMPFIND algorithm, and turn out to be in larger regions of self-gravitating gas. And here we arrive at what may be the most tantalizing point of all. A debate that has enlivened star-formation theory for nearly a decade is how stars acquire their mass<sup>5</sup>. Do they accrete their gas from relatively isolated cores, or do

they accrete material as they move about in a broader gravitational potential, gathering mass through competition with other dense regions in the same gravitationally bound region<sup>6</sup> (the 'competitive accretion' picture)? Although neither hypothesis is amenable to definitive observational tests, the dendrogram method developed by Goodman *et al.* has the potential to answer this question and to identify the real conditions in which stars form. ■

Ralph E. Pudritz is in the Department of Physics and Astronomy, McMaster University, 1280 Main

Street West, Hamilton, Ontario L8S 4M1, Canada.  
e-mail: pudritz@physics.mcmaster.ca

1. Goodman, A. A. *et al.* *Nature* **457**, 63–66 (2009).
2. Bate, M. R. *Mon. Not. R. Astron. Soc.* (in the press); preprint at <http://arxiv.org/abs/0811.0163> (2008).
3. Motte, F., André, P. & Neri, R. *Astron. Astrophys.* **336**, 150–172 (1998).
4. Johnstone, D. *et al.* *Astrophys. J.* **545**, 327–339 (2000).
5. McKee, C. F. & Ostriker, E. C. *Annu. Rev. Astron. Astrophys.* **45**, 565–687 (2007).
6. Bonnell, I. A. & Bate, M. R. *Mon. Not. R. Astron. Soc.* **370**, 488–494 (2006).

## GAME THEORY

# How to treat those of ill repute

Bettina Rockenbach and Manfred Milinski

**A much-needed theoretical analysis deals with whether the principle known as 'costly punishment' helps to maintain cooperation in human society. It will prompt a fresh wave of experiments and theory.**

Human societies are built on cooperation, especially on reciprocation<sup>1</sup> — I help you and you help me, or I help you and someone else helps me. In the first case, help is directly reciprocated by help. In the second, called indirect reciprocity, I gain a good reputation and so I can expect help when in need.

But how shall I treat someone with a bad reputation? Shall I just refuse help or shall I punish this person at a cost to myself? Costly punishment can enhance cooperation<sup>2,3</sup> in experiments with human subjects, but potentially with no net benefit<sup>4</sup>: the costs of punishment usually, although not always<sup>5</sup>, neutralize gains from enhanced cooperation. On page 79 of this issue, Ohtsuki *et al.*<sup>6</sup> describe a theoretical test of whether either refusing help to or punishing someone with a bad reputation might lead to a cooperative society. They conclude that, except under certain rare conditions, punishment does not produce that outcome.

When you meet someone needing help, you can help (cooperate), refuse to help (defect) or not only refuse to help but, in addition, decrease the needy person's wealth (punish). Both cooperation and punishment are costly for you, but respectively create a larger benefit or larger loss for the person needing help. Defection is cost neutral.

How you behave depends on the reputation — good or bad — of the needy person, and depends upon your 'action rule'. An example is 'cooperate with someone with a good reputation and defect with someone with a bad reputation' (CD). The reputation you yourself gain by applying your action rule depends on the social norm of your society. Under the norm 'stern-judging', for example, you gain a good reputation when cooperating with good or when defecting with bad, and a bad reputation

in all other cases. Thus CD always leads to a good reputation under stern-judging.

Another action rule, CP, prescribes 'cooperate with good and punish bad'. Under stern-judging, with CP you will achieve a good reputation when you interact with someone with a good reputation and a bad reputation when you interact with someone with a bad reputation. But under a different social norm, 'shunning' (cooperation with good or punishment of bad leads to a good reputation), CP will always provide you with a good reputation (Fig. 1).

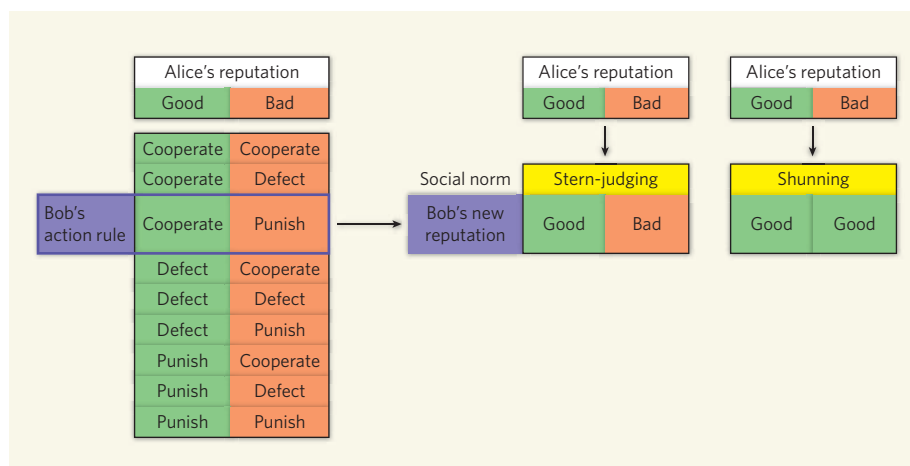
In their simplest model, Ohtsuki *et al.*<sup>6</sup> assume that everybody has the same opinion of the reputation of another person or has the same level of fallibility in assigning an

incorrect reputation. For such a society, they test for each of the 64 different social norms (Fig. 1) whether an action rule exists that both generates a cooperative society and is evolutionarily stable — meaning one that resists replacement (invasion) by any of the other eight possible action rules (Fig. 1).

Ohtsuki *et al.* find that the two action rules that induce cooperation and resist invasion are those described above — CD under stern-judging and CP under shunning. Nonetheless, the average pay-off is lower if the action rule uses costly punishment, while the stability conditions are less restrictive.

However, which parameters determine which rule is most efficient in the sense of leading to the highest average pay-off at equilibrium? It turns out that a crucial one is the accuracy of assigning the correct reputation to everybody. If this accuracy is too low then only a DD action rule is efficient, under which nobody cooperates. If the accuracy is high enough, then CD can be efficient. For intermediate values of the accuracy, there is a small window in which CP can be efficient, as reflected in the title of the paper<sup>6</sup>: "Indirect reciprocity provides only a narrow margin of efficiency for costly punishment."

In a further step in their modelling, Ohtsuki *et al.*<sup>6</sup> dropped the assumption that all good or bad reputations are publicly known, and allowed individual knowledge of reputations. They found that the stability of both CD and CP is lost when there is the smallest error in distinguishing between good and bad. When individuals start to communicate with each other and adjust their assessments of everybody's reputation, the CP action rule can be stably maintained. Then, when reputations become even more publicly agreed upon through more efficient gossip, both CD and CP are stably maintained under their



**Figure 1 | Action rules, social norms and the story of Alice and Bob.** Bob meets Alice and learns whether Alice has a good or a bad reputation. On the basis of that, Bob may either help (cooperate), refuse help (defect) or refuse help and, in addition, decrease Alice's wealth (punish). How Bob reacts is specified in his action rule. Bob's own reputation depends on how society's social norm evaluates Bob's reaction to Alice's reputation. The social norm 'stern-judging' evaluates cooperation with good as good and punishment of bad as bad; the social norm 'shunning' evaluates cooperation with good as good and punishment of bad as good. Each social norm specifies which reputation to assign for each of the 6 possible scenarios (3 actions of Bob for 2 reputations of Alice). This leads to  $2^6 = 64$  social norms.



they accrete material as they move about in a broader gravitational potential, gathering mass through competition with other dense regions in the same gravitationally bound region<sup>6</sup> (the 'competitive accretion' picture)? Although neither hypothesis is amenable to definitive observational tests, the dendrogram method developed by Goodman *et al.* has the potential to answer this question and to identify the real conditions in which stars form. ■

Ralph E. Pudritz is in the Department of Physics and Astronomy, McMaster University, 1280 Main

Street West, Hamilton, Ontario L8S 4M1, Canada.  
e-mail: pudritz@physics.mcmaster.ca

1. Goodman, A. A. *et al.* *Nature* **457**, 63–66 (2009).
2. Bate, M. R. *Mon. Not. R. Astron. Soc.* (in the press); preprint at <http://arxiv.org/abs/0811.0163> (2008).
3. Motte, F., André, P. & Neri, R. *Astron. Astrophys.* **336**, 150–172 (1998).
4. Johnstone, D. *et al.* *Astrophys. J.* **545**, 327–339 (2000).
5. McKee, C. F. & Ostriker, E. C. *Annu. Rev. Astron. Astrophys.* **45**, 565–687 (2007).
6. Bonnell, I. A. & Bate, M. R. *Mon. Not. R. Astron. Soc.* **370**, 488–494 (2006).

## GAME THEORY

# How to treat those of ill repute

Bettina Rockenbach and Manfred Milinski

**A much-needed theoretical analysis deals with whether the principle known as 'costly punishment' helps to maintain cooperation in human society. It will prompt a fresh wave of experiments and theory.**

Human societies are built on cooperation, especially on reciprocation<sup>1</sup> — I help you and you help me, or I help you and someone else helps me. In the first case, help is directly reciprocated by help. In the second, called indirect reciprocity, I gain a good reputation and so I can expect help when in need.

But how shall I treat someone with a bad reputation? Shall I just refuse help or shall I punish this person at a cost to myself? Costly punishment can enhance cooperation<sup>2,3</sup> in experiments with human subjects, but potentially with no net benefit<sup>4</sup>: the costs of punishment usually, although not always<sup>5</sup>, neutralize gains from enhanced cooperation. On page 79 of this issue, Ohtsuki *et al.*<sup>6</sup> describe a theoretical test of whether either refusing help to or punishing someone with a bad reputation might lead to a cooperative society. They conclude that, except under certain rare conditions, punishment does not produce that outcome.

When you meet someone needing help, you can help (cooperate), refuse to help (defect) or not only refuse to help but, in addition, decrease the needy person's wealth (punish). Both cooperation and punishment are costly for you, but respectively create a larger benefit or larger loss for the person needing help. Defection is cost neutral.

How you behave depends on the reputation — good or bad — of the needy person, and depends upon your 'action rule'. An example is 'cooperate with someone with a good reputation and defect with someone with a bad reputation' (CD). The reputation you yourself gain by applying your action rule depends on the social norm of your society. Under the norm 'stern-judging', for example, you gain a good reputation when cooperating with good or when defecting with bad, and a bad reputation

in all other cases. Thus CD always leads to a good reputation under stern-judging.

Another action rule, CP, prescribes 'cooperate with good and punish bad'. Under stern-judging, with CP you will achieve a good reputation when you interact with someone with a good reputation and a bad reputation when you interact with someone with a bad reputation. But under a different social norm, 'shunning' (cooperation with good or punishment of bad leads to a good reputation), CP will always provide you with a good reputation (Fig. 1).

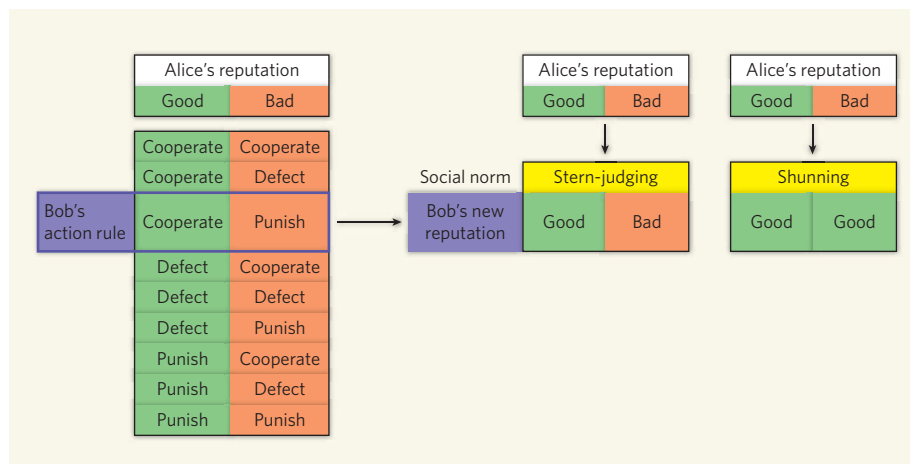
In their simplest model, Ohtsuki *et al.*<sup>6</sup> assume that everybody has the same opinion of the reputation of another person or has the same level of fallibility in assigning an

incorrect reputation. For such a society, they test for each of the 64 different social norms (Fig. 1) whether an action rule exists that both generates a cooperative society and is evolutionarily stable — meaning one that resists replacement (invasion) by any of the other eight possible action rules (Fig. 1).

Ohtsuki *et al.* find that the two action rules that induce cooperation and resist invasion are those described above — CD under stern-judging and CP under shunning. Nonetheless, the average pay-off is lower if the action rule uses costly punishment, while the stability conditions are less restrictive.

However, which parameters determine which rule is most efficient in the sense of leading to the highest average pay-off at equilibrium? It turns out that a crucial one is the accuracy of assigning the correct reputation to everybody. If this accuracy is too low then only a DD action rule is efficient, under which nobody cooperates. If the accuracy is high enough, then CD can be efficient. For intermediate values of the accuracy, there is a small window in which CP can be efficient, as reflected in the title of the paper<sup>6</sup>: "Indirect reciprocity provides only a narrow margin of efficiency for costly punishment."

In a further step in their modelling, Ohtsuki *et al.*<sup>6</sup> dropped the assumption that all good or bad reputations are publicly known, and allowed individual knowledge of reputations. They found that the stability of both CD and CP is lost when there is the smallest error in distinguishing between good and bad. When individuals start to communicate with each other and adjust their assessments of everybody's reputation, the CP action rule can be stably maintained. Then, when reputations become even more publicly agreed upon through more efficient gossip, both CD and CP are stably maintained under their



**Figure 1 | Action rules, social norms and the story of Alice and Bob.** Bob meets Alice and learns whether Alice has a good or a bad reputation. On the basis of that, Bob may either help (cooperate), refuse help (defect) or refuse help and, in addition, decrease Alice's wealth (punish). How Bob reacts is specified in his action rule. Bob's own reputation depends on how society's social norm evaluates Bob's reaction to Alice's reputation. The social norm 'stern-judging' evaluates cooperation with good as good and punishment of bad as bad; the social norm 'shunning' evaluates cooperation with good as good and punishment of bad as good. Each social norm specifies which reputation to assign for each of the 6 possible scenarios (3 actions of Bob for 2 reputations of Alice). This leads to  $2^6 = 64$  social norms.

corresponding social norms. Experimental studies<sup>7</sup> in indirect reciprocity have shown that gossip can indeed serve as a surrogate for direct observation. But it will take further empirical research to find out whether gossip is efficient enough to re-establish both CD and CP.

The next question addressed by Ohtsuki *et al.* was which kind of society someone might prefer to live in. To this end, they simulated one society with CD under stern-judging and another with CP under shunning. When individuals can choose freely between them, the CD society — that with the higher expected pay-off — is preferred. Thus, the CP rule loses to CD when people can choose between societies with different norms.

This last result can be compared with our own experimental work with human subjects<sup>8</sup>. We found that when individuals had the choice between a CD-only society and one with both CD and CP, they ultimately preferred the latter. Compared with a CP-only control, punishing acts were largely reduced in the CD plus CP society but were concentrated on the most uncooperative players, rendering

them more cooperative. Our experimental societies in which CD and CP coexisted were more efficient than those with only CP, suggesting that they have a more complex action rule: respond to good with cooperate, to bad with defect, and to very bad with punish. Such a possibility sets a challenge for theorists.

Finally, given that Ohtsuki *et al.* show that the social norm of a society determines which action will prevail, another task is to uncover the social norms of real societies and analyse which action rule to expect. Ohtsuki *et al.* assume that all social norms are equally likely. However, the more information a norm requires in order to develop, the more susceptible it is to errors and the more costly is the information acquisition<sup>9</sup>. Such restrictions may challenge any social norm that otherwise dominates: for example, in an experimental study<sup>10</sup>, the subjects had a majority social norm similar to shunning that was 'low observation' and 'memory demanding'.

Ultimately, study of the joint evolution of social norms and action rules under natural constraints is the goal for the future.

Ohtsuki *et al.* have prepared the ground for that endeavour. ■

Bettina Rockenbach is in the Department of Economics, University of Erfurt, Nordhäuser Straße 63, D-99089 Erfurt, Germany. Manfred Milinski is in the Department of Evolutionary Ecology, Max Planck Institute for Evolutionary Biology, August-Thienemann-Straße 2, D-24306 Plön, Germany.

e-mails: [bettina.rockenbach@uni-erfurt.de](mailto:bettina.rockenbach@uni-erfurt.de); [milinski@evolbio.mpg.de](mailto:milinski@evolbio.mpg.de)

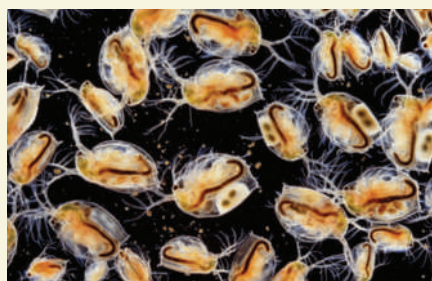
1. Nowak, M. A. & Sigmund, K. *Nature* **437**, 1291–1298 (2005).
2. Fehr, E. & Gächter, S. *Nature* **415**, 137–140 (2002).
3. Gülerk, Ö., Irlenbusch, B. & Rockenbach, B. *Science* **312**, 108–111 (2006).
4. Dreber, A., Rand, D. G., Fudenberg, D. & Nowak, M. A. *Nature* **452**, 348–351 (2008).
5. Gächter, S., Renner, E. & Sefton, M. *Science* **322**, 1510 (2008).
6. Ohtsuki, H., Iwasa, Y. & Nowak, M. A. *Nature* **457**, 79–82 (2009).
7. Sommerfeld, R. D., Krambeck, H.-J., Semmann, D. & Milinski, M. *Proc. Natl Acad. Sci. USA* **104**, 17435–17440 (2007).
8. Rockenbach, B. & Milinski, M. *Nature* **444**, 718–723 (2006).
9. Brandt, H. & Sigmund, K. *Proc. Natl Acad. Sci. USA* **102**, 2666–2670 (2005).
10. Milinski, M., Semmann, D., Bakker, T. C. M. & Krambeck, H.-J. *Proc. R. Soc. Lond. B* **268**, 2495–2501 (2001).

## DARWIN 200

# A natural selection



As Charles Darwin showed so convincingly, the fauna of islands provide excellent subjects for investigating evolution. The creature on the left, *Anolis sagrei*, is a case in point. Jonathan Losos and colleagues studied this lizard in experimental work, carried out in 2003, that involved the introduction of a predator of this species onto six islands in the Bahamas. Six other islands acted as controls. Losos and colleagues' aim was



to test the hypothesis that, when organisms experience new environments, behavioural change prevents the operation of natural selection (they concluded that in this case it did not).

Readers can find out for themselves what the authors did at [www.nature.com/evolutiongems](http://www.nature.com/evolutiongems). The paper concerned is one of "Fifteen evolutionary gems: A resource for those wishing to spread awareness of evolution by natural

selection", which from today will be available as a collection on the *Nature* website. The "gems" are all papers, published in *Nature* over the past decade or so, that demonstrate the enduring explanatory power of Darwinian natural selection.

Examples are drawn from the fossil record, from extant organisms in natural and experimental habitats, and from molecular studies. The other images here — of (clockwise from top left) water fleas, *Daphnia*

*magna*; fledglings of the great tit, *Parus major*; sticklebacks, *Gasterosteus aculeatus*; and a garter snake of the species *Thamnophis sirtalis* — provide a taster of the subjects of other papers in the collection.

The papers are free to download and disseminate, and each is accompanied by a brief editorial introduction to its context and significance. ■

See also Editorial, page 8.



corresponding social norms. Experimental studies<sup>7</sup> in indirect reciprocity have shown that gossip can indeed serve as a surrogate for direct observation. But it will take further empirical research to find out whether gossip is efficient enough to re-establish both CD and CP.

The next question addressed by Ohtsuki *et al.* was which kind of society someone might prefer to live in. To this end, they simulated one society with CD under stern-judging and another with CP under shunning. When individuals can choose freely between them, the CD society — that with the higher expected pay-off — is preferred. Thus, the CP rule loses to CD when people can choose between societies with different norms.

This last result can be compared with our own experimental work with human subjects<sup>8</sup>. We found that when individuals had the choice between a CD-only society and one with both CD and CP, they ultimately preferred the latter. Compared with a CP-only control, punishing acts were largely reduced in the CD plus CP society but were concentrated on the most uncooperative players, rendering

them more cooperative. Our experimental societies in which CD and CP coexisted were more efficient than those with only CP, suggesting that they have a more complex action rule: respond to good with cooperate, to bad with defect, and to very bad with punish. Such a possibility sets a challenge for theorists.

Finally, given that Ohtsuki *et al.* show that the social norm of a society determines which action will prevail, another task is to uncover the social norms of real societies and analyse which action rule to expect. Ohtsuki *et al.* assume that all social norms are equally likely. However, the more information a norm requires in order to develop, the more susceptible it is to errors and the more costly is the information acquisition<sup>9</sup>. Such restrictions may challenge any social norm that otherwise dominates: for example, in an experimental study<sup>10</sup>, the subjects had a majority social norm similar to shunning that was 'low observation' and 'memory demanding'.

Ultimately, study of the joint evolution of social norms and action rules under natural constraints is the goal for the future.

Ohtsuki *et al.* have prepared the ground for that endeavour. ■

Bettina Rockenbach is in the Department of Economics, University of Erfurt, Nordhäuser Straße 63, D-99089 Erfurt, Germany. Manfred Milinski is in the Department of Evolutionary Ecology, Max Planck Institute for Evolutionary Biology, August-Thienemann-Straße 2, D-24306 Plön, Germany.

e-mails: bettina.rockenbach@uni-erfurt.de; milinski@evolbio.mpg.de

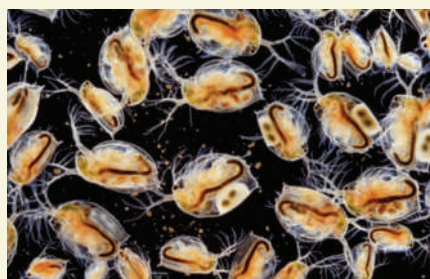
1. Nowak, M. A. & Sigmund, K. *Nature* **437**, 1291–1298 (2005).
2. Fehr, E. & Gächter, S. *Nature* **415**, 137–140 (2002).
3. Gülerk, Ö., Irlenbusch, B. & Rockenbach, B. *Science* **312**, 108–111 (2006).
4. Dreber, A., Rand, D. G., Fudenberg, D. & Nowak, M. A. *Nature* **452**, 348–351 (2008).
5. Gächter, S., Renner, E. & Sefton, M. *Science* **322**, 1510 (2008).
6. Ohtsuki, H., Iwasa, Y. & Nowak, M. A. *Nature* **457**, 79–82 (2009).
7. Sommerfeld, R. D., Krambeck, H.-J., Semmann, D. & Milinski, M. *Proc. Natl Acad. Sci. USA* **104**, 17435–17440 (2007).
8. Rockenbach, B. & Milinski, M. *Nature* **444**, 718–723 (2006).
9. Brandt, H. & Sigmund, K. *Proc. Natl Acad. Sci. USA* **102**, 2666–2670 (2005).
10. Milinski, M., Semmann, D., Bakker, T. C. M. & Krambeck, H.-J. *Proc. R. Soc. Lond. B* **268**, 2495–2501 (2001).

## DARWIN 200

# A natural selection



As Charles Darwin showed so convincingly, the fauna of islands provide excellent subjects for investigating evolution. The creature on the left, *Anolis sagrei*, is a case in point. Jonathan Losos and colleagues studied this lizard in experimental work, carried out in 2003, that involved the introduction of a predator of this species onto six islands in the Bahamas. Six other islands acted as controls. Losos and colleagues' aim was



to test the hypothesis that, when organisms experience new environments, behavioural change prevents the operation of natural selection (they concluded that in this case it did not).

Readers can find out for themselves what the authors did at [www.nature.com/evolutiongems](http://www.nature.com/evolutiongems). The paper concerned is one of "Fifteen evolutionary gems: A resource for those wishing to spread awareness of evolution by natural

selection", which from today will be available as a collection on the *Nature* website. The "gems" are all papers, published in *Nature* over the past decade or so, that demonstrate the enduring explanatory power of Darwinian natural selection.

Examples are drawn from the fossil record, from extant organisms in natural and experimental habitats, and from molecular studies. The other images here — of (clockwise from top left) water fleas, *Daphnia*

*magna*; fledglings of the great tit, *Parus major*; sticklebacks, *Gasterosteus aculeatus*; and a garter snake of the species *Thamnophis sirtalis* — provide a taster of the subjects of other papers in the collection.

The papers are free to download and disseminate, and each is accompanied by a brief editorial introduction to its context and significance. ■

See also Editorial, page 8.

## REVIEWS

# 18 years of science with the Hubble Space Telescope

Julianne J. Dalcanton<sup>1,2</sup>

**After several decades of planning, the Hubble Space Telescope (HST) was launched in 1990 as the first of NASA's Great Observatories. After a rocky start arising from an error in the fabrication of its main mirror, it went on to change forever many fields of astronomy, and to capture the public's imagination with its images. An ongoing programme of servicing missions has kept the telescope on the cutting edge of astronomical research. Here I review the advances made possible by the HST over the past 18 years.**

**T**he impact of the HST on the public imagination is large, yet many are surprised to learn that the HST is rather modestly sized in comparison with modern telescopes (Fig. 1). Its mirror is only a generous arm span across, with less than  $\sim 1/15$ th the light-gathering area of the largest telescopes available on the ground. Its instruments are likewise far from state of the art, containing detectors whose technologies are sometimes more than a decade old. What then accounts for the HST's tremendous scientific and public impact?

The HST's successes can be attributed to the same three key factors influencing real estate: location, location, location. Raising the HST above most of the Earth's atmosphere has allowed it to escape a host of problems that limit telescopes on the ground. The first of the

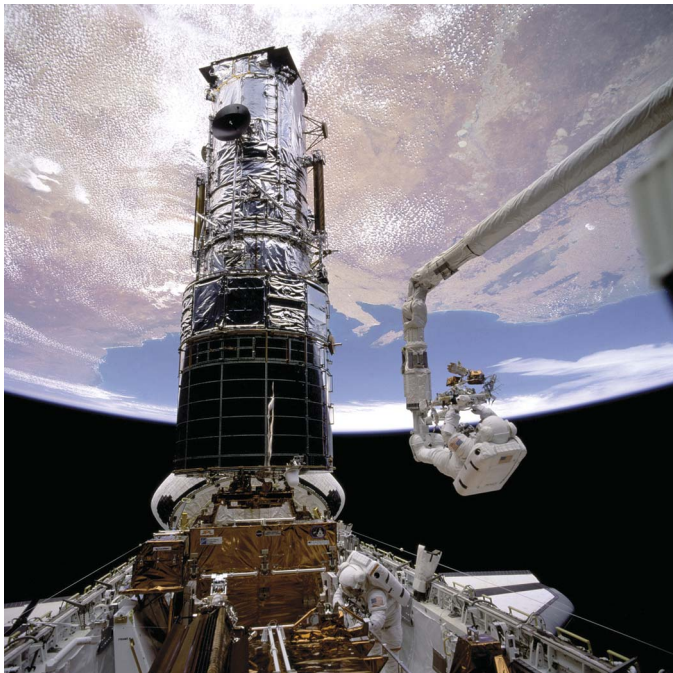
resulting improvements is the dramatic increase in a telescope's resolution—the smallest angular separation that can be reliably detected (Fig. 2). Images taken with the HST can distinguish features that are separated by less than a tenth of an arc second (comparable to the angle spanned by half a millimetre when seen from a kilometre away). Comparable images made from the ground are typically blurred by a factor of ten relative to those made with the HST. Moreover, because the HST's view of the Universe is unperturbed by the turbulent, chaotic atmosphere, the images and spectra taken by the telescope are stable and reproducible. This stability allows for an unprecedented level of precision when measuring the brightness and structure of astrophysical objects.

The second, less publicized benefit of the HST's location is the darkness of space-based imaging. From the ground, the night sky is not actually all that dark. Atoms in the upper atmosphere absorb energy from sunlight during the day, and then re-radiate that energy as light during the night. In much the same way that you can see more stars from a dark campground than from a bright city street, from space the HST can measure astronomical features fainter than those ground-based telescopes can detect against the glowing background of the night sky. This increase in contrast is especially pronounced at ultraviolet wavelengths, where photons are otherwise absorbed by the Earth's atmosphere and are thus inaccessible from the ground. Access to this portion of the electromagnetic spectrum is critical for tracing the abundance of individual elements in astrophysical gases, and for interpreting observations of galaxies at large distances.

In this article I summarize some of the many areas where the HST has dramatically changed our understanding of the Universe. These include tracing the structure and evolution of stars from their births through to their deaths, demonstrating the pervasiveness of black holes and their links to galaxy formation, tracking the evolution of galaxies over billions of years, and testing fundamental models of the expansion of the Universe.

## Distance scale

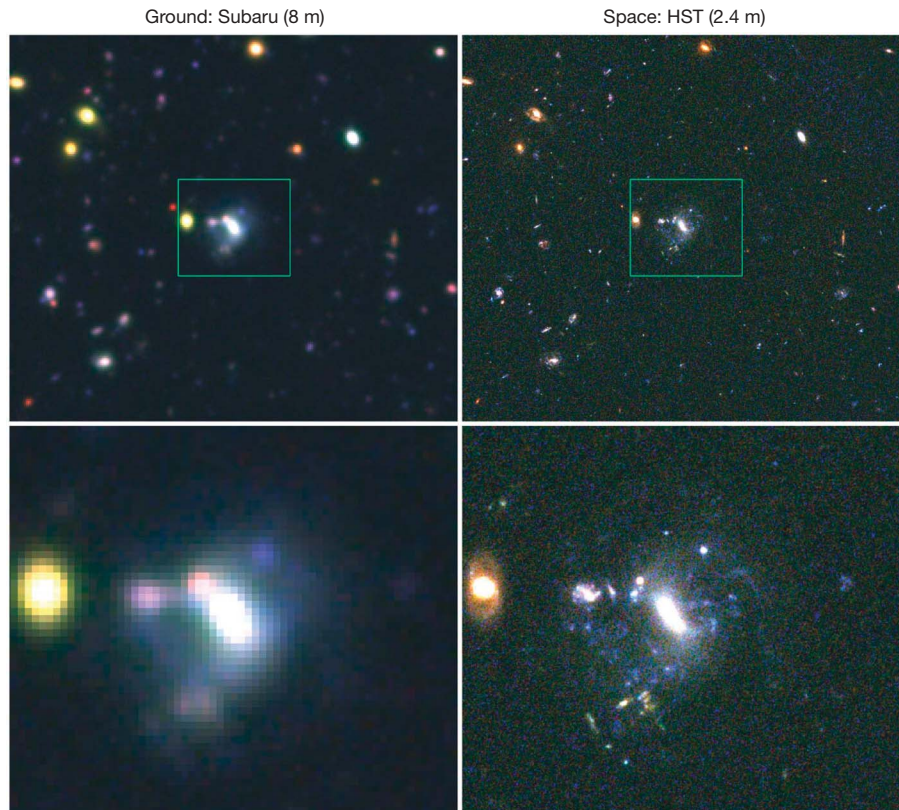
One of the most consistently vexing issues facing astronomers is measurement of the distances to astronomical objects. It perhaps seems odd that something so basic, and so easily judged on Earth, can be so difficult in an astronomical context. However, in astronomy we observe systems whose energy outputs and distances can differ by a dozen orders of magnitude. We therefore often have little a priori knowledge with which to estimate the basic size and energy



**Figure 1 | The Hubble Space Telescope, while docked with the Space Shuttle Endeavour during servicing.** The new instruments installed during such missions have kept HST at the cutting edge of astronomy. Image courtesy of NASA.

<sup>1</sup>University of Washington, Seattle, Washington 98195, USA. <sup>2</sup>Max-Planck-Institut für Astronomie, D-69117 Heidelberg, Germany.





**Figure 2 | Images taken from space have superb spatial resolution.** Images of a distant galaxy taken from the ground (with the Subaru 8-m telescope (National Astronomical Observatory of Japan) on Mauna Kea) are blurred

in comparison with images of the same galaxy made with the HST. Images courtesy of NASA.

scales of a given system. Are we observing a faint object that happens to be close, or a bright object that is much farther away? Moreover, even when we can make reasonable assumptions about an object's distance, we frequently lack the precision needed for critical cosmological tests.

Over the past century, astronomers have built up an elaborate distance 'ladder', which uses the distances to the nearest stars to infer the distances to more distant objects. We start the ladder with stars that are sufficiently close that their position on the sky shifts during the Earth's orbit around the Sun. This parallax effect (which you can see by alternately shutting your left and right eyes) can be used to calculate a distance on the basis of the principles of geometry only, making it robust against systematic errors. The only limit to the method is the precision with which we can measure slight angular shifts. With higher resolution, parallax measurements can be pushed to much larger distances.

The HST excels at parallax measurements owing to both its superb resolution and its Fine Guidance Sensor (FGS). Although the FGS is primarily responsible for keeping the telescope oriented precisely during scientific observations, it can also be used to make milli-arc-second measurements of the angular positions of individual stars. Moreover, the FGS is quite sensitive and can measure accurate parallaxes for stars that are much fainter than those that could be studied with the earlier parallax mission satellite HIPPARCOS (the High Precision Parallax Collecting Satellite). By measuring parallaxes of fainter stars, the HST can probe to larger distances, and sample a larger volume of the Galaxy. This allows the HST to calibrate the luminosities of truly rare stars which are difficult to find nearby, but that are crucial for establishing more distant rungs on the distance ladder.

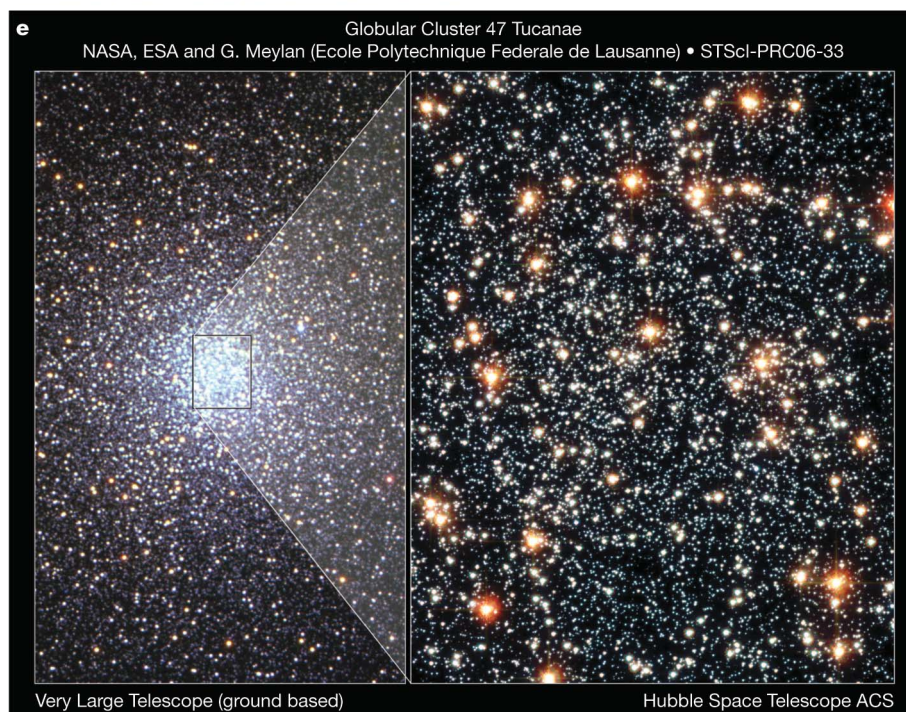
One of the most important types of such stars are Cepheids, a class of pulsating stars that are up to thousands of times more luminous than the Sun, and can thus be identified in galaxies far outside the

Milky Way owing to their brightnesses and characteristic periodic variations in luminosity. A Cepheid's luminosity is tightly correlated with the duration of its pulsation cycle. Once this correlation is calibrated, one can then deduce the distance to a Cepheid by measuring the duration of its pulsation, inferring its luminosity and then comparing the luminosity with the apparent brightness. The accuracy of the derived distance is typically limited by the accuracy of the initial period–luminosity relationship. The FGS on the HST has made direct parallax measurements of this relationship<sup>1</sup>, greatly reducing the uncertainties in Cepheid-based distances. Although this period–luminosity relationship may potentially vary with the metallicity of the galaxy<sup>2</sup>, with appropriate corrections it can reliably be used for many massive spiral galaxies. Previously, climbing from reliable parallax measurements of nearby stars to Cepheid-based distances to spiral galaxies required many complicated steps up the distance ladder. With the new HST results, astronomers can make this step in a single jump.

Identifying Cepheids in nearby galaxies requires the HST's resolution. Although Cepheids are luminous, they can appear to be blended together with the light from many other stars in images from ground-based telescopes, making it impossible to identify the luminosity fluctuations or the brightness of the one star in the cluster that happens to be a Cepheid. In contrast, the high angular resolution of

**Figure 3 | A montage showing the life cycle of stars.** **a**, The Orion nebula is a dense cloud of gas within which new stars are forming. The inset shows a close-up of a young star surrounded by a dusty disk that may eventually form planets. **b**, Young blue stars surrounded by leftover natal gas in the Large Magellanic Cloud, a nearby dwarf galaxy. **c**, The evolving star V838 Monocerotis, whose past flashes have illuminated material that was ejected in an early stellar wind. **d**, A planetary nebula forms as a dying star sheds its outer layers, leaving behind a fading white dwarf at its centre. **e**, A globular cluster containing many old stars. The faintest stars in the cluster are white dwarfs. Images courtesy of NASA.







the HST allows Cepheids to be separated from their neighbouring stars out to vastly larger distances, increasing the number of galaxies with potentially reliable distances by a factor of close to 1,000. Establishing Cepheid-based distances to nearby galaxies was one of the Key Projects identified before the HST's launch. The Cepheid Key Project has since used hundreds of orbits to analyse Cepheids in 36 galaxies, all of which were used to calibrate even more distant rungs of the ladder (such as supernovae and the Tully–Fisher relationship).

These Cepheid distances have become the bedrock foundation of our understanding of the size scale of the distant Universe. Distances to galaxies can be inferred from the 'Hubble diagram'—a linear relationship between the distance to a galaxy and the speed at which it appears to be receding from us. The slope of this relationship is known as the Hubble constant, and once its value is known it is possible to measure recessional velocities to galaxies (which is straightforward) and use them to infer distances (which is difficult). The HST's observations of Cepheids have been essential in establishing the distance portion of the Hubble diagram, and thus have laid the foundations for modern measurements of the Hubble constant. Before the HST, the Hubble constant was known only to within a factor of two, but with the new Cepheid observations the uncertainty in the distance scale of the Universe has dropped to roughly 10%.

### The life cycle of stars

Although stars are frequently assumed to be constant and unchanging features of the firmament, they are in fact evolving dynamic systems. New stars condense out of gaseous nebulae, and old stars evolve through planetary nebulae and supernovae into white dwarfs, neutron stars and black holes. These processes—star formation and evolution—are critical to understanding many features of the Universe, including the evolution of galaxies, the dispersal of chemical elements and the distribution and energetics of gas.

Some of the HST's most lasting (and beautiful) contributions to stellar astronomy have been its studies of star-forming regions like the Orion nebula (Fig. 3). In these regions, luminous massive stars ionize the gas cloud from which they coalesced, causing the cloud to glow brightly in various emission lines. The HST's earliest observations of the Orion nebula<sup>3,4</sup> revealed that it was peppered with a remarkable population of young stars surrounded by dense disks of gas and dust. These disks are undoubtedly remnants of the late accretion phase during which the stars condensed. Although the presence of such disks had been inferred from theory<sup>5</sup> and from observations with the Very Large Array<sup>6</sup>, the HST's superior image resolution revealed the first true pictures of the disks' structures and physical properties.

The observed 'proto-planetary disks' are exactly the type of system that astronomers expect will evolve into planetary systems like those hosted by our own Sun. The ubiquity of the disks seen in the HST images gave strong hints that our Solar System is unlikely to be unique, and recent studies have indeed found direct evidence for the 'extrasolar' planetary systems that are expected to evolve from the disks seen in the images. The HST is studying intermediate stages of this process directly, by mapping the detailed structures of disks that have lost their gas but are still dominated by dust that will eventually be either incorporated into planets or dispersed<sup>7–10</sup>. Although individual cases had been observed from the ground, the HST has proven the ubiquity of such systems. Moreover, with its superb resolution, the HST is uncovering evidence of structure in these debris disks, indicating that they have been shaped by the gravitational influence of planets orbiting the young, distant host stars<sup>11</sup>.

In addition to revealing the properties of young stars, the HST continues to help us understand the older stars that dominate the stellar content of our Galaxy. Although many of these stars are relatively ordinary stars like the Sun, the majority are much fainter and are of lower mass. At these low masses, the stars begin to have more in common with planets, hosting complex atmospheres and sometimes lacking a central energy source (brown dwarfs). Luckily, many of

these stars are also locked into gravitationally bound orbiting pairs, known as binaries, which serve as laboratories for measuring the properties of stars. The orbits of binary stars can be used to derive their masses, and, because the stars in the binary can be assumed to have formed simultaneously, it can also be assumed that they have a common age and metallicity. Binaries are thus particularly useful when studying poorly understood types of star. Unfortunately, most stars are far enough away that the images of the two stars in the binary cannot be separated in a typical image, making it difficult to measure the orbits, luminosities, colours and spectra of the two components. This step is particularly difficult when one of the stars is much fainter than the other, as is common for low-mass stars, which are among the most complex and least well-understood of normal stars. However, with the HST's spatial resolution, image stability and low background, the properties of binary stars can easily be measured. The HST has thus powered critical advances in our understanding of low-mass stars and brown dwarfs, by identifying the earliest brown dwarf binaries<sup>12,13</sup>, characterizing the relationship between mass and luminosity for low-mass stars<sup>14</sup>, making dynamical mass measurements for field brown dwarfs<sup>15,16</sup>, finding unusual T dwarf binaries<sup>17</sup> and constraining the rate of binarity at the lowest stellar masses<sup>18,19</sup>. It has also been an essential tool in simply quantifying the numbers of the lowest mass stars and failed brown dwarfs, both in young<sup>20,21</sup> and old<sup>22,23</sup> stellar clusters.

The HST is yielding surprises even for more ordinary stars. For most of the past century, astronomers believed that the simplest collections of stars could be found in globular clusters, which are dense, gravitationally bound groups of  $\sim 10^4$ – $10^6$  stars. The stars in globular clusters were thought to have formed at exactly the same time, from a single cloud of gas, giving them identical ages and chemical compositions. Unfortunately, our ability to measure the properties of individual stars in globular clusters is compromised by the high stellar densities within the clusters, and only with the HST can individual stars be separated from one another in an image. Recently, the exceptional photometric stability and high resolution of the HST have revealed that some globular clusters have far more complex properties than were previously thought possible: some host stars that formed in multiple events and/or with distinct chemical compositions<sup>24,25</sup>. Although this work is new, it has already begun overturning half a century of thought on the nature of these supposedly 'simple' systems.

### The death of stars

After their life of quiescent hydrogen burning is over, stars undergo dramatic changes. When the primary energy source at the centre of a star becomes exhausted, the star is forced to undergo dramatic evolution. During this evolution, the star expands drastically, sheds much of its mass and either explodes or fades away, leaving behind a bizarre stellar remnant (such as a white dwarf, neutron star or black hole).

These late phases of stellar evolution are surprisingly brief on typical astronomical timescales. Astronomers have used the HST to track the final stages of mass loss in rapidly evolving stars like  $\eta$  Carinae, using time-resolved measurements to watch the expansion of the stellar ejecta in real time<sup>26–28</sup>. They have also revealed the ghosts of past mass ejection from evolving stars like V838 Monocerotis (ref. 29) or from SN 1987A (refs 30–32) using 'light echoes' that have bounced off expanding shells of matter ejected farther and farther into the recent past. Equally striking are the HST images of planetary nebulae<sup>33</sup>, formed after most mass loss is complete in lower mass stars. Like the images of  $\eta$  Carinae and V838 Monocerotis, the planetary nebulae show a rich level of complexity<sup>34–36</sup>, some of which varies over year-long timescales<sup>37,38</sup>. Most importantly, in almost none of these cases does the distribution of ejected mass look spherical, in spite of the nearly spherical shape of the initial star. This asymmetry suggests that the process of mass loss in stars is complex, and involves 'shaping' of the stellar winds by rotation, magnetic fields and/or orbiting stellar companions.

Similar evidence that even apparently symmetric stars have asymmetric deaths comes from HST identifications of neutron stars, which are incredibly dense stellar remnants left behind after the supernova explosion of a massive star. Although they begin their lives as stars larger than the Sun, neutron stars are only a few kilometres across, making them very faint at optical wavelengths. The optical emission provides a critical means of constraining their temperatures and sizes, which in turn tells a great deal about the strange nuclear matter within the stars. The HST has been able to identify the optical counterparts of several neutron stars, and through repeated observations has found that some are hurtling through the galaxy at a surprisingly high speed of more than  $100 \text{ km s}^{-1}$  (ref. 39). These observations suggest that the supernova somehow gives the dense stellar remnant a tremendous 'kick', shooting it out of the supernova, despite the initial star being essentially spherical.

The HST is also shedding light on the connection between supernovae from dying stars and mysterious  $\gamma$ -ray bursts. At the HST's launch, it was not known if these fast bursts of intense  $\gamma$ -ray radiation were from sources inside our galaxy or far outside it. Subsequent satellites showed that they were occurring outside our galaxy, and HST observations of the fading optical transients further localized some of the bursts to individual galaxies, in regions containing young, massive stars<sup>40–42</sup>. HST thus convincingly showed that these highly energetic events are tied to the very last phases of the lives of massive stars. The theoretical models resulting from these observations are suggesting that such 'hypernovae' may be essential for establishing the unusual elemental abundance patterns seen in the oldest stars.

### Black holes

The same explosions that lead to supernovae and  $\gamma$ -ray bursts can also lead to the formation of remnant black holes. These enigmatic objects form when mass collapses to within an extremely small radius, known as the Schwarzschild radius. For a star like the Sun to turn into a black hole, it would need to shrink by a factor of more than 200,000, down to a radius of less than 3 km. When a massive star runs out of fuel it collapses to within its Schwarzschild radius, becoming sufficiently dense that even light cannot escape from the tremendous gravitation field, and leading to the name 'black hole'. In spite of their name, black holes can be responsible for some of the most luminous astrophysical phenomena, owing to the high temperatures of the material that they sometimes accrete. However, in many cases black holes are truly dark, and their presence can only be inferred from their gravitational influence on the material around them.

Key to finding a dark black hole is to look for evidence of a large amount of mass being in a very small space. We can deduce that mass is present by the motions produced by the gravitational pull of the black hole. When these motions are very fast, and show rapid changes in velocity and/or direction over a very small distance, then we can infer the presence of a black hole. However, we need measurements of velocities that probe exceptionally small physical distances to ensure that the measured mass is in a small enough volume that it can be explained only by a black hole. Finding black holes in even the nearest galaxies thus requires the exceptional angular resolution of the HST.

Over the past several decades, the HST has revolutionized our view of black holes and the role they have in galaxy formation. At the launch of the HST, it was generally assumed that supermassive black holes were needed to explain the unusually high luminosities and peculiar spectra of distant quasi-stellar objects (QSOs) and lower luminosity nearby active galactic nuclei. However, there was then no firm evidence that QSOs were always hosted by galaxies, or that all galaxies currently hosted remnant black holes left over from an early epoch when QSOs were common. Some ground-based observations strongly suggested that a handful of the nearest galaxies might host dense, massive objects at their centres, but the identification of these as black holes was far from confirmed<sup>43</sup>.

Since its launch, the HST has been engaged in a systematic observational assault on these issues. One of the earliest HST Key Projects was responsible for establishing that all QSOs (which were thought to be powered by black holes) were indeed hosted by galaxies<sup>44</sup>. Later, dynamical modelling of HST spectra in nearby galaxies confirmed the presence of black holes through their gravitational influence on the surrounding stars<sup>45–48</sup>. These studies led to dozens of robust measurements of central black hole masses, which then revealed an extremely tight correlation between black hole mass and the mass of the galactic bulge (the smooth, rounded stellar component at the galaxy's centre)<sup>49,50</sup>; this relationship had been hinted at in previous ground-based work, but with much larger uncertainty and scatter. Moreover, the modelling showed that even dormant galaxies without luminous active nuclei—a sign of accretion onto a black hole—also hosted black holes, confirming that central, massive black holes are a generic feature of galaxy formation.

This surprising result has yet to be understood. The black holes that form during the deaths of massive stars have relatively small masses, and are unlikely to be more than 20 times more massive than the Sun. By contrast, black holes at the centres of galaxies are millions or billions of times more massive than the Sun, leaving open questions about how the supermassive black holes formed, how they migrated to the centre and how their masses 'knew' about the masses of their host galaxies. HST observations are beginning to unravel some of these questions, hinting that intermediate,  $\sim 10^4$ -solar-mass, black holes might exist in some globular clusters<sup>51,52</sup> that could later accrete into the galactic centre, but that some bulgeless galaxies (like our neighbour M33) have no evidence for a central black hole at all<sup>53</sup>. The HST's resolution has also been essential to the observation of structures that may be either analogues or precursors of central black holes. These 'nuclear clusters' are dense stellar knots found at the centres of some galaxies. The masses of nuclear clusters are also correlated with galaxy properties<sup>54–56</sup>, raising the suspicion that they are potentially visible manifestations of the processes that may lead to supermassive black holes. The HST has shown that at least some of these clusters are disk-like, suggesting that gas accretion may have a role in the growth of massive structures at the centres of galaxies<sup>57</sup>.

Given the presence of massive black holes at the centres of most galaxies, it might be expected that the black holes have helped to shape the distribution of stars around them. Such an effect should be strongest in massive galaxies, owing to the more massive black holes within. Indeed, some of the earliest investigations with the HST revealed strong differences between the central stellar distributions of massive and low-mass galaxies<sup>58,59</sup>; these differences, although recently questioned<sup>60</sup>, have been confirmed in many subsequent studies<sup>61–63</sup>. Galaxies were found to have a sharp bimodality at their centres, with the more massive galaxies having flat 'cores' in their light distributions and lower mass galaxies having a bright central cusp. Some models predict that these flat cores could be produced by pairs of central black holes, the orbits of which can fling stars out of the galactic centres<sup>64</sup> either before or during merging. Although the observed trend can be interpreted as a response to the presence of central black holes, there are many other properties of the galaxies that similarly correlate with the central light profile, such as the overall shape of the galaxy (box-like versus disk-like, triaxial versus axisymmetric) and the internal kinetics (rotating versus non-rotating). These trends may indicate that the dichotomy between core and cusp galaxies may transcend the properties of the central black holes, and may instead be intimately linked with the formation of the galaxy as a whole.

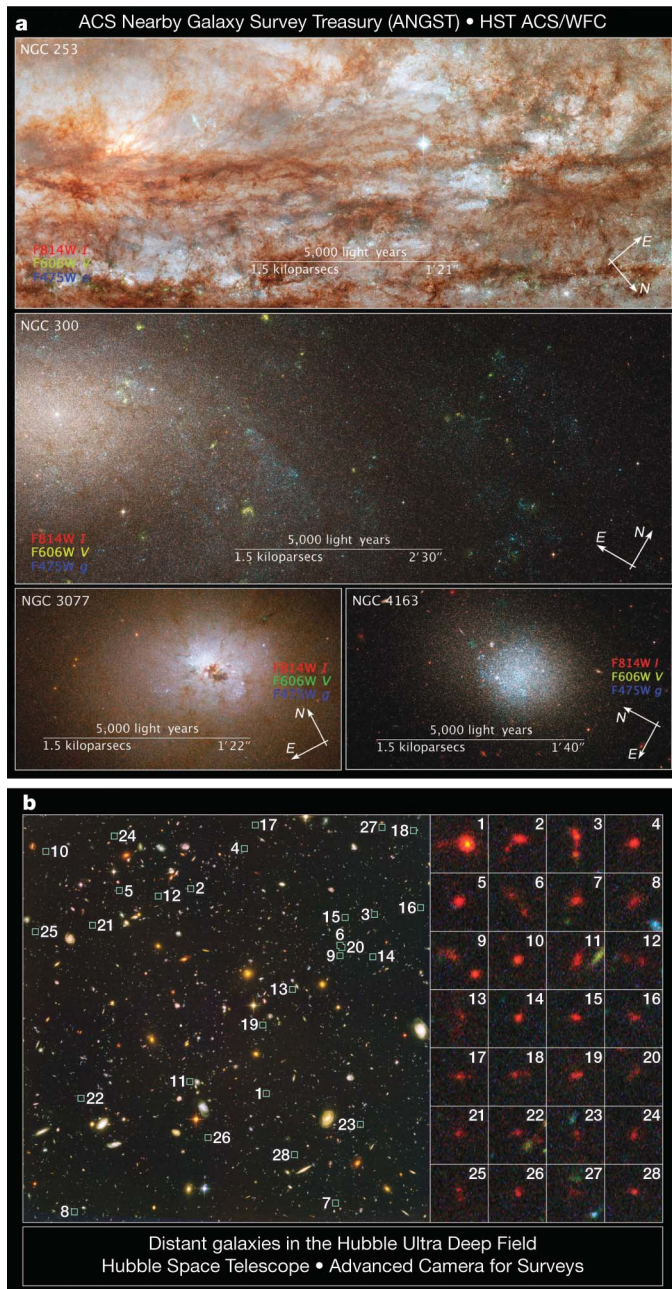
### Growth of galaxies

At the time that the HST was launched, astronomers had long been used to seeing the familiar, regular forms of galaxies at the present day. However, there were long-standing questions as to how these forms came about, and how they may have changed and evolved through time. In principle, the question could be answered by looking at distant



galaxies, which are so far away that light emitted many billions of years ago is only now reaching us. Thus, in progressively more distant galaxies we can see earlier and earlier snapshots of the galaxies' evolution, reaching back to a time when the Universe was only a fraction of its present age—an effect much like receiving postcards sent from ever more distant places, carrying old news from the time they were sent. Unfortunately, owing to blurring produced by the Earth's atmosphere, these distant galaxies appeared only as unresolved smudges in images taken from the ground.

With the launch of the HST, astronomers got their first look at the shapes and structures of young galaxies, and what they saw was surprising (Fig. 4). Instead of looking like close analogues of the spiral and elliptical galaxies we see today, images of the distant Universe revealed a large population of galaxies that looked not unlike insects splattered on a windscreen. This population can be seen most dramatically in images of the Hubble Deep Fields<sup>65,66</sup> and the later-generation Ultra Deep Field<sup>67</sup>, made with a more sensitive, larger-format camera.



**Figure 4 | Galaxies near and far.** **a**, Images showing the diverse properties of galaxies. **b**, Unusually shaped distant galaxies in the Hubble Ultra Deep Field. Images courtesy of NASA.

The HST deep field images revealed that young galaxies were built out of fragments, through a process that we now understand to be the hierarchical build-up of smaller galaxies into ever larger ones. The HST images were key in establishing the current galaxy formation model, in which galaxy properties are shaped by the continual merger and accretion of matter over cosmic timescales<sup>68–74</sup>. Parallel work in clusters of galaxies showed the influence of merging on larger scales, using the combination of HST images and spectroscopy<sup>75–82</sup>. Only with the HST could we measure the complicated structures and sizes<sup>83,84</sup> of these younger evolving galaxies.

Also key in establishing this change in model was the huge investment made in measuring relative distances to the galaxies in the Hubble Deep Field (HDF). Measuring relative distances requires time-consuming spectroscopy of faint individual galaxies with which one can measure recessional velocities. Although such programmes are expensive in terms of limited telescope resources, given the unprecedented allocation of the HST's resources in producing the HDF, the astronomical community responded with an equally unprecedented investment of spectroscopic observations of the HDF galaxies<sup>85–87</sup>. These spectroscopic observations became key in establishing the link between a galaxy's observed colour and its distance. This 'photometric redshift' technique had been developed for nearby galaxies<sup>88</sup>, but only with the HDF was it confirmed to be a viable method for measuring distances to the most distant galaxies<sup>89</sup>. Since the verification of this technique in the HDF, there has been an explosion in the use of photometric redshifts, allowing major surveys<sup>90–93</sup> to skip time-consuming spectroscopy when establishing the relative distances of galaxies. Photometric redshifts also allow astronomers to estimate distances to galaxies that are otherwise too faint for spectroscopic observations. Some of the earliest work applying photometric redshifts were used for the HDF itself<sup>94–96</sup>, and showed most clearly the assembly of galaxies from small lumps into large coherent units, and the accompanying rise in the overall rate of star formation in the Universe<sup>94,97</sup>.

In addition to the great strides made through its *in situ* studies of distant young galaxies, the HST has made equal progress using the complementary tool of studying the fossil record that galaxy formation has left behind. This record is embedded in the individual stars that formed during a galaxy's lifetime. Seen from the ground, these billions of stars blur into a single smooth image. However, when viewed at the highest spatial resolution, the smooth galaxy image breaks into millions of individual points, revealing individual stars, much like the pointillism seen in a Seurat painting. The colours and brightnesses of these stars encode information about their ages and chemical compositions<sup>98</sup>, and allow us to infer the past history of star formation in the galaxy.

With the accurate measurements provided by the HST's spatial resolution and photometric stability, astronomers have characterized the stellar populations of the oldest extended structures in galaxies<sup>99–106</sup>, tracing the earliest epochs of a galaxy's assembly. They have also used the HST to map the ebb and flow of star formation across the disk of nearby galaxies, tracing the detailed evolution of star formation over hundreds of millions of years<sup>107</sup>.

## Cosmology

Some of the most widely recognized successes of the HST have been in the field of cosmology. Cosmology attempts to measure and understand the most basic large-scale features of the Universe, such as its age, composition and size. Constraining these quantities relies primarily on astronomical observations, which are among the only tools that can access the distances and ages over which the Universe's evolution and structure can be measured.

The age of the Universe can be inferred in multiple ways. The most secure measurements of the age come from the oldest stars we see today, whose ages place a secure lower limit on the age of the Universe. To derive ages for a population of stars, we ideally want to find a cluster of stars that were all born at the same time. The stars within the cluster all evolve at different rates, such that the most

massive stars quickly become supernovae, the intermediate-mass stars (like the Sun) evolve into planetary nebulae and leave behind a faint remnant white dwarf, and the lowest mass stars last for the age of the Universe. The age of the stellar cluster can then be inferred from the masses of the stars that are left behind.

There has been an elegant series of HST measurements using this technique, focusing on white dwarfs. After a white dwarf forms, it no longer has a source of energy in its core. It therefore cools steadily, dropping in temperature and brightness at a rate that can be calculated from theoretical models. These models can be compared with the observed distribution of brightnesses to infer the age of the white dwarf, placing a robust lower limit on the age of the star<sup>108–112</sup>. Unfortunately, white dwarfs are sufficiently faint that they can be difficult to isolate from other stars. In a clever twist, astronomers have used repeated, deep observations with the HST to separate faint cluster stars from intervening and background objects<sup>113–115</sup>. The stars in the cluster all share a common velocity, causing them to show slight shifts in position between the two epochs of observation. These shifts are distinct from those shown by foreground and background stars, allowing the cluster stars to be isolated down to very faint luminosities with little confusion. The resulting measurements of the distribution of white dwarf brightnesses are tremendously precise, and have placed secure lower limits on the age of the Universe that do not depend on uncertain observations at high redshift or assumptions about the underlying cosmology. Instead, they rely solely on our understanding of the physics of stars, making this a superb independent check on a fundamental cosmological measurement. This measurement could not have been made without the high resolution and the photometric (and astrometric) stability of the HST.

An independent probe of the Universe's age comes from the HST's measurement of the Hubble constant using Cepheid distances. In addition to providing a fundamental distance scale for the Universe, the value of the Hubble constant is directly related to the rate at which the Universe is expanding at the present day. Its value is therefore critical in extrapolating between the present and the Big Bang, and thus in establishing the duration of the expansion.

Whereas the Hubble constant constrains the rate of the Universe's expansion at the present day, astronomical observations can also be used to constrain the Universe's past expansion. Our understanding of the past rate of expansion has been revolutionized by the study of the brightnesses of high-redshift supernovae. Progressively more distant supernovae should be increasingly fainter, in a way that depends sensitively on the exact model of the Universe in which we live. Measurements of distant supernovae have shown that they appear unusually faint<sup>116–122</sup>. However, the typical luminosities of supernovae are well understood, and their apparent faintness must thus indicate that they are much farther away than had been anticipated in the early 1990s. For this to be true, the Universe must be expanding far faster than in any model that includes only relatively normal forms of matter. This evidence for 'dark energy' or 'cosmological acceleration' was one of the ground-breaking scientific discoveries of the late 1990s.

The HST has been essential in discovering and monitoring ever more distant supernovae, which probe early epochs in the Universe's expansion. At such large distances, it is difficult to measure the properties of supernovae from the ground, owing to their faintness and the blurring of their light with the surrounding galaxy. With the HST, however, the supernovae can be separated cleanly from the host galaxy, allowing the brightnesses to be measured. These precisely measured brightnesses constrain the distances to the supernovae, and limit the nature of the expansion at these early times, providing a strong limit on possible models for the nature of the dark energy.

Supernova measurements of the dark energy suggest that it dominates the mass-energy density of the Universe, but there is still an enormous quantity of dark matter pervading space. Although its true nature eludes us, its mass and distribution can be probed through indirect techniques. One of the most successful techniques is through

the distortion that dark matter produces on images of background galaxies. Like looking through the bottom of a wine glass, the mass of dark matter distorts light passing through it, producing a warped image of all the images seen through the dark matter. Normally, this effect is too small to be measured, but for the densest concentrations of dark matter, found in massive clusters of galaxies, the effect can be dramatic, producing enormous arcs and ring-like structures around the clusters<sup>123</sup>. These images are among the most spectacular extragalactic images produced by HST.

With the precision offered by the HST's resolution, astronomers have mastered techniques that use the observed distortions to reconstruct the distribution of mass that produced them. This makes it possible to map the dark matter distribution, in spite of not being able to see it directly<sup>124–126</sup>. One of the most striking uses of this technique can be seen in recent observations of two colliding galaxy clusters (the 'bullet cluster'). Maps of the dark matter constructed from HST images show a remarkable separation of the dark matter from the gas in the cluster<sup>127</sup>. Like two colliding streams of water, the gas was essentially stopped in the collision, whereas the dark matter and the galaxies passed through. These observations are striking confirmation that the dark matter is not associated with the gas (which dominates the normal matter in the clusters), and that the dark matter acts like 'collisionless' particles that interact only through gravity and not any electromagnetic process. Thus, in spite of not knowing what the dark matter is, the HST has allowed us to learn what it is not, and how it behaves on large scales.

## Gas in the Universe

Although cosmology tells us that the majority of the mass in the Universe is hidden in the dark sector, nearly all the light we see comes from normal, familiar forms of matter. This matter is only a small fraction of the Universe's mass, but understanding what forms it takes and how it is distributed through space constrains many astrophysical questions.

The majority of normal baryonic matter is thought to be in the gas phase, and can produce both emission lines and absorption lines in astrophysical spectra. The location (in wavelength) and amplitude of these lines can be used to constrain the chemical composition, temperature, density and pressure of the gas. Although optical spectroscopy from the ground can routinely study some of the spectroscopic features produced by astrophysical gases, many of the most useful features are found at ultraviolet wavelengths, and can only be studied from space.

Ultraviolet spectroscopy with the HST has allowed astronomers to track where gas is and how the amount of matter in the gas phase has changed with time. One of the earliest HST Key Projects mapped the distribution of gas in the (relatively) nearby Universe using the absorption lines produced by hydrogen along the line of sight to more distant quasars<sup>128</sup>. Gas had previously been tracked in this manner, but only at larger distances that probe the properties of the Universe long ago. The evolution of the gas phase could therefore not be tracked continuously to the present, leaving the final destination of the gas uncertain. Moreover, when restricted to such large distances, little was known about exactly what astronomical objects the gas was associated with. Was it bound to galaxies? Or was it distributed in the otherwise empty volume of space between galaxies? Only with the HST's ultraviolet sensitivity could the amount and distribution of this gas be tracked to the present day, through the local distribution of galaxies. The HST thus made it possible to follow the evolution of the primary gas reservoir in the Universe from the earliest times to the present<sup>129–132</sup> and, moreover, to match the gas seen in absorption to other nearby galaxies<sup>133–138</sup>. These early studies showed that there was rapid evolution in the amount of broadly distributed gas, as it was absorbed into galaxies, and that there were multiple populations of absorbing clouds, some that were clearly associated with galaxies and others that resided in the 'cosmic web'. More recently, ultraviolet spectroscopy with the HST has



uncovered an even larger reservoir of gas<sup>139</sup>, which simulations<sup>140</sup> and subsequent ultraviolet observations with NASA's Far Ultraviolet Spectroscopic Explorer<sup>141,142</sup> suggest is hiding in a hot, diffuse plasma surrounding nearby galaxies. This ionized gas may contain even more mass than is currently found in stars and cool gas.

In addition to tracking the broad movement of gas from the cosmic web into galaxies, the HST has been essential in tracing the detailed properties of gas within individual galaxies, and particularly within the Milky Way. Ultraviolet spectroscopy of nearby stars shows a rich network of absorption lines produced by several dozen different elements found in the gas in our galaxy. By analysing the depths of various absorption features, astronomers have shown in detail how different elements move in and out of the gas phase as they are expelled by stars and/or condense into solid dust grains<sup>143</sup>. Tracking these changes improves our understanding both of how these elements are produced deep within stars and of the properties of the dust which clouds almost every astronomical observation.

### A new way forward for astronomy

Beyond the direct scientific impact of its observations, the HST has had a significant role in shaping the culture of astronomy. In many scientific fields, data are proprietary, and are owned indefinitely by the groups who performed an individual experiment. Ground-based astronomy was no different in this regard. The HST, however, was recognized to be a sufficiently unique and limited resource as to make an indefinite proprietary period unjustifiable. Instead, any data taken by the HST are open for anyone to use after only one year. This policy is an effective multiplier for the HST's impact, allowing all data to be used by multiple groups, for multiple purposes, facilitated by the HST's excellent archiving programme.

This movement towards non-proprietary data was greatly accelerated by the HDF, which was imaged using an unprecedented allocation of observing time donated by the director of the Space Telescope Science Institute, Bob Williams. The images were reduced by staff at the institute and released to the public immediately. The wider astronomical community responded by investing their own resources in spectroscopy and follow-up imaging at other wavelengths, using other facilities. This coming together of the community to generate a shared, non-proprietary data set was essentially unprecedented, but has since become the model for the majority of large astronomical projects. Almost all major astronomical surveys are now proposed with the expectation that the data and data products will be publicly released during the project, rather than held in perpetuity by those few who instigated the programme. This new mode of operating has democratized astronomy by opening astronomical research to scientists that are at relatively under-resourced institutions, allowing researchers at small colleges, or in poor countries, to have access to some of the finest data sets in the world. The policies established by the HST's oversight board were crucial steps in ushering in this cultural change.

### The future of the HST

Owing to the HST's upcoming servicing mission, the installation's future looks as bright as its past. With increased imaging sensitivity at ultraviolet and near-infrared wavelengths, enhanced ultraviolet spectroscopy and the restoration of several failed instruments, the HST will surpass any capability it had in the past. Even as astronomers master the art of achieving high-resolution imaging from the ground through adaptive optics, the HST's photometric stability, ultraviolet sensitivity and low background will remain unparalleled. Unfortunately, when the HST is finally decommissioned, the window of high-resolution optical and ultraviolet imaging will be sadly closed for decades to come.

1. Benedict, G. F. *et al.* Hubble Space Telescope fine guidance sensor parallaxes of Galactic Cepheid Variable stars: Period-luminosity relations. *Astron. J.* **133**, 1810–1827 (2007).

2. Kennicutt, R. C. Jr *et al.* The Hubble Space Telescope Key Project on the extragalactic distance scale. XIII. The metallicity dependence of the Cepheid distance scale. *Astrophys. J.* **498**, 181–194 (1998).
3. O'Dell, C. R., Wen, Z. & Hu, X. Discovery of new objects in the Orion nebula on HST images - Shocks, compact sources, and protoplanetary disks. *Astrophys. J.* **410**, 696–700 (1993).
4. O'Dell, C. R. & Wen, Z. Postrefurbishment mission Hubble Space Telescope images of the core of the Orion Nebula: Proplyds, Herbig-Haro objects, and measurements of a circumstellar disk. *Astrophys. J.* **436**, 194–202 (1994).
5. Shu, F. H., Adams, F. C. & Lizano, S. Star formation in molecular clouds - Observation and theory. *Annu. Rev. Astron. Astrophys.* **25**, 23–81 (1987).
6. Felli, M., Churchwell, E., Wilson, T. L. & Taylor, G. B. The radio continuum morphology of the Orion Nebula - From 10 arcmin to 0.1 arcsec resolution. *Astron. Astrophys. Suppl. Ser.* **98**, 137–164 (1993).
7. Schneider, G. *et al.* NICMOS imaging of the HR 4796A circumstellar disk. *Astrophys. J.* **513**, L127–L130 (1999).
8. Heap, S. R. *et al.* Space Telescope Imaging Spectrograph coronagraphic observations of  $\beta$  Pictoris. *Astrophys. J.* **539**, 435–444 (2000).
9. Krist, J. E. *et al.* Hubble Space Telescope Advanced Camera for Surveys coronagraphic imaging of the AU Microscopii debris disk. *Astron. J.* **129**, 1008–1017 (2005).
10. Golimowski, D. A. *et al.* Hubble Space Telescope ACS multiband coronagraphic imaging of the debris disk around  $\beta$  Pictoris. *Astron. J.* **131**, 3109–3130 (2006).
11. Kalas, P., Graham, J. R. & Clampin, M. A planetary system as the origin of structure in Fomalhaut's dust belt. *Nature* **435**, 1067–1070 (2005).
12. Martin, E. L., Brandner, W. & Basri, G. A search for companions to nearby brown dwarfs: The binary DENIS-P J1228.2-1547. *Science* **283**, 1718–1720 (1999).
13. Lowrance, P. J. *et al.* A candidate substellar companion to CD -33 deg 7795 (TWA 5). *Astrophys. J.* **512**, L69–L72 (1999).
14. Henry, T. J. *et al.* The optical mass-luminosity relation at the end of the main sequence (0.08 - 0.20  $M_{\odot}$ ). *Astrophys. J.* **512**, 864–873 (1999).
15. Bouy, H. *et al.* First determination of the dynamical mass of a binary L dwarf. *Astron. Astrophys.* **423**, 341–352 (2004).
16. Golimowski, D. A., Burrows, C. J., Kulkarni, S. R., Oppenheimer, B. R. & Bruckardt, R. A. Wide Field Planetary Camera 2 observations of the brown dwarf Gliese 229B: Optical colors and orbital motion. *Astron. J.* **115**, 2579–2586 (1998).
17. Burgasser, A. J. *et al.* Binarity in brown dwarfs: T dwarf binaries discovered with the Hubble Space Telescope Wide Field Planetary Camera 2. *Astrophys. J.* **586**, 512–526 (2003).
18. Martin, E. L. *et al.* Membership and multiplicity among very low mass stars and brown dwarfs in the Pleiades cluster. *Astrophys. J.* **543**, 299–312 (2000).
19. Allen, P. R. Star formation via the little guy: A Bayesian study of ultracool dwarf imaging surveys for companions. *Astrophys. J.* **668**, 492–506 (2007).
20. Luhman, K. L. *et al.* The initial mass function of low-mass stars and brown dwarfs in young clusters. *Astrophys. J.* **540**, 1016–1040 (2000).
21. Najita, J. R., Tiede, G. P. & Carr, J. S. From stars to superplanets: The low-mass initial mass function in the young cluster IC 348. *Astrophys. J.* **541**, 977–1003 (2000).
22. King, I. R., Anderson, J., Cool, A. M. & Piotto, G. The luminosity function of the globular cluster NGC 6397 near the limit of hydrogen burning. *Astrophys. J.* **492**, L37–L40 (1998).
23. Richer, H. B. *et al.* Deep Advanced Camera for Surveys imaging in the globular cluster NGC 6397: the cluster color-magnitude diagram and luminosity function. *Astron. J.* **135**, 2141–2154 (2008).
24. Bedin, L. R. *et al.*  $\omega$  Centauri: The population puzzle goes deeper. *Astrophys. J.* **605**, L125–L128 (2004).
25. Piotto, G. *et al.* A triple main sequence in the globular cluster NGC 2808. *Astrophys. J.* **661**, L53–L56 (2007).
26. Currie, D. G. *et al.* Astrometric analysis of the homunculus of eta Carinae with the Hubble Space Telescope. *Astron. J.* **112**, 1115–1127 (1996).
27. Morse, J. A. *et al.* Hubble Space Telescope proper-motion measurements of the  $\eta$  Carinae Nebula. *Astrophys. J.* **548**, L207–L211 (2001).
28. O'Dell, C. R. & Doi, T. High proper motion features in the central Orion nebula. *Astron. J.* **125**, 277–287 (2003).
29. Bond, H. E. *et al.* An energetic stellar outburst accompanied by circumstellar light echoes. *Nature* **422**, 405–408 (2003).
30. Jakobsen, P. *et al.* First results from the Faint Object Camera - SN 1987A. *Astrophys. J.* **369**, L63–L66 (1991).
31. Burrows, C. J. *et al.* Hubble Space Telescope observations of the SN 1987A triple ring nebula. *Astrophys. J.* **452**, 680–684 (1995).
32. Plait, P. C., Lundqvist, P., Chevalier, R. A. & Kirshner, R. P. HST observations of the ring around SN 1987A. *Astrophys. J.* **439**, 730–751 (1995).
33. Balick, B. & Frank, A. Shapes and shaping of planetary nebulae. *Annu. Rev. Astron. Astrophys.* **40**, 439–486 (2002).
34. Sahai, R. & Trauger, J. T. Multipolar bubbles and jets in low-excitation planetary nebulae: Toward a new understanding of the formation and shaping of planetary nebulae. *Astron. J.* **116**, 1357–1366 (1998).
35. Balick, B. *et al.* FLIERs and other microstructures in planetary nebulae. IV. Images of elliptical PNs from the Hubble Space Telescope. *Astron. J.* **116**, 360–371 (1998).
36. O'Dell, C. R., Balick, B., Hajian, A. R., Henney, W. J. & Burkert, A. Knots in nearby planetary nebulae. *Astron. J.* **123**, 3329–3347 (2002).
37. Reed, D. S. *et al.* Hubble Space Telescope measurements of the expansion of NGC 6543: Parallax distance and nebular evolution. *Astron. J.* **118**, 2430–2441 (1999).

38. Fernández, R., Monteiro, H. & Schwarz, H. E. Proper motion and kinematics of the ansae in NGC 7009. *Astrophys. J.* **603**, 595–598 (2004).
39. Walter, F. M. The proper motion, parallax, and origin of the isolated neutron star RX J185635–3754. *Astrophys. J.* **549**, 433–440 (2001).
40. Fruchter, A. S. *et al.* Hubble Space Telescope and Palomar imaging of GRB 990123: Implications for the nature of gamma-ray bursts and their hosts. *Astrophys. J.* **519**, L13–L16 (1999).
41. Bloom, J. S., Kulkarni, S. R. & Djorgovski, S. G. The observed offset distribution of gamma-ray bursts from their host galaxies: A robust clue to the nature of the progenitors. *Astron. J.* **123**, 1111–1148 (2002).
42. Fruchter, A. S. *et al.* Long  $\gamma$ -ray bursts and core-collapse supernovae have different environments. *Nature* **441**, 463–468 (2006).
43. Kormendy, J. & Richstone, D. Inward bound? The search for supermassive black holes in galactic nuclei. *Annu. Rev. Astron. Astrophys.* **33**, 581–624 (1995).
44. Bahcall, J. N., Kirhakos, S., Saxe, D. H. & Schneider, D. P. Hubble Space Telescope images of a sample of 20 nearby luminous quasars. *Astrophys. J.* **479**, 642–658 (1997).
45. Ferrarese, L., Ford, H. C. & Jaffe, W. Evidence for a massive black hole in the active galaxy NGC 4261 from Hubble Space Telescope images and spectra. *Astrophys. J.* **470**, 444–459 (1996).
46. Kormendy, J. *et al.* Hubble Space Telescope spectroscopic evidence for a  $2 \times 10^9 M_{\odot}$  black hole in NGC 3115. *Astrophys. J.* **459**, L57–L60 (1996).
47. van der Marel, R. P., Cretton, N., de Zeeuw, P. T. & Rix, H.-W. Improved evidence for a black hole in M32 from HST/FOS Spectra. II. Axisymmetric dynamical models. *Astrophys. J.* **493**, 613–631 (1998).
48. Gebhardt, K. *et al.* Axisymmetric, three-integral models of galaxies: A massive black hole in NGC 3379. *Astron. J.* **119**, 1157–1171 (2000).
49. Ferrarese, L. & Merritt, D. A fundamental relation between supermassive black holes and their host galaxies. *Astrophys. J.* **539**, L9–L12 (2000).
50. Gebhardt, K. *et al.* A relationship between nuclear black hole mass and galaxy velocity dispersion. *Astrophys. J.* **539**, L13–L16 (2000).
51. Gerssen, J. *et al.* Hubble Space Telescope evidence for an intermediate-mass black hole in the globular cluster M15. II. Kinematic analysis and dynamical modeling. *Astron. J.* **124**, 3270–3288 (2002).
52. Gebhardt, K., Rich, R. M. & Ho, L. C. A  $20,000 M_{\odot}$  black hole in the stellar cluster G1. *Astrophys. J.* **578**, L41–L45 (2002).
53. Gebhardt, K. *et al.* M33: A galaxy with no supermassive black hole. *Astron. J.* **122**, 2469–2476 (2001).
54. Ferrarese, L. *et al.* A fundamental relation between compact stellar nuclei, supermassive black holes, and their host galaxies. *Astrophys. J.* **644**, L21–L24 (2006).
55. Wehner, E. H. & Harris, W. E. From supermassive black holes to dwarf elliptical nuclei: A mass continuum. *Astrophys. J.* **644**, L17–L20 (2006).
56. Balcells, M., Graham, A. W. & Peletier, R. F. Galactic bulges from Hubble Space Telescope NICMOS Observations: Central Galaxian objects, and nuclear profile slopes. *Astrophys. J.* **665**, 1084–1103 (2007).
57. Seth, A. C., Dalcanton, J. J., Hodge, P. W. & Debattista, V. P. Clues to nuclear star cluster formation from edge-on spirals. *Astron. J.* **132**, 2539–2555 (2006).
58. Lauer, T. R. *et al.* The centers of early-type galaxies with HST. I. An observational survey. *Astron. J.* **110**, 2622–2654 (1995).
59. Faber, S. M. *et al.* The centers of early-type galaxies with HST. IV. Central parameter relations. *Astron. J.* **114**, 1771–1796 (1997).
60. Ferrarese, L. *et al.* The ACS Virgo cluster survey. VI. Isothermal analysis and the structure of early-type galaxies. *Astrophys. J. Suppl. Ser.* **164**, 334–434 (2006).
61. Rest, A. *et al.* WFPC2 images of the central regions of early-type galaxies. I. The data. *Astron. J.* **121**, 2431–2482 (2001).
62. Ravindranath, S., Ho, L. C., Peng, C. Y., Filippenko, A. V. & Sargent, W. L. W. Central structural parameters of early-type galaxies as viewed with Nicmos on the Hubble Space Telescope. *Astron. J.* **122**, 653–678 (2001).
63. Lauer, T. R. *et al.* The centers of early-type galaxies with Hubble Space Telescope. VI. Bimodal central surface brightness profiles. *Astrophys. J.* **664**, 226–256 (2007).
64. Begelman, M. C., Blandford, R. D. & Rees, M. J. Massive black hole binaries in active galactic nuclei. *Nature* **287**, 307–309 (1980).
65. Williams, R. E. *et al.* The Hubble Deep Field: Observations, data reduction, and galaxy photometry. *Astron. J.* **112**, 1335–1389 (1996).
66. Williams, R. E. *et al.* The Hubble Deep Field South: Formulation of the observing campaign. *Astron. J.* **120**, 2735–2746 (2000).
67. Beckwith, S. V. W. *et al.* The Hubble Ultra Deep Field. *Astron. J.* **132**, 1729–1755 (2006).
68. Glazebrook, K., Ellis, R., Santiago, B. & Griffiths, R. The morphological identification of the rapidly evolving population of faint galaxies. *Mon. Not. R. Astron. Soc.* **275**, L19–L22 (1995).
69. Abraham, R. G. *et al.* Galaxy morphology to  $I=25$  mag in the Hubble Deep Field. *Mon. Not. R. Astron. Soc.* **279**, L47–L52 (1996).
70. Odewahn, S. C., Windhorst, R. A., Driver, S. P. & Keel, W. C. Automated morphological classification in deep Hubble Space Telescope UVBI fields: Rapidly and passively evolving faint galaxy populations. *Astrophys. J.* **472**, L13–L16 (1996).
71. Driver, S. P., Windhorst, R. A. & Griffiths, R. E. The contribution of late-type/irregulars to the faint galaxy counts from Hubble Space Telescope Medium-Deep Survey images. *Astrophys. J.* **453**, 48–64 (1995).
72. Giallisco, M., Steidel, C. C. & Macchetto, F. D. Hubble Space Telescope imaging of star-forming galaxies at redshifts  $Z > 3$ . *Astrophys. J.* **470**, 189–194 (1996).
73. van Dokkum, P. G., Franx, M., Kelson, D. D. & Illingworth, G. D. Luminosity evolution of early-type galaxies to  $Z = 0.83$ : Constraints on formation epoch and Omega. *Astrophys. J.* **504**, L17–L21 (1998).
74. Le Fèvre, O. *et al.* Hubble Space Telescope imaging of the CFRS and LDSS redshift surveys - IV. Influence of mergers in the evolution of faint field galaxies from  $z \sim 1$ . *Mon. Not. R. Astron. Soc.* **311**, 565–575 (2000).
75. Dressler, A., Oemler, A. J., Butcher, H. R. & Gunn, J. E. The morphology of distant cluster galaxies. 1: HST observations of CL 0939+4713. *Astrophys. J.* **430**, 107–120 (1994).
76. Dressler, A. *et al.* Evolution since  $Z = 0.5$  of the morphology-density relation for clusters of galaxies. *Astrophys. J.* **490**, 577–591 (1997).
77. Ellis, R. S. *et al.* The homogeneity of spheroidal populations in distant clusters. *Astrophys. J.* **483**, 582–596 (1997).
78. Stanford, S. A., Eisenhardt, P. R. & Dickinson, M. The evolution of early-type galaxies in distant clusters. *Astrophys. J.* **492**, 461–479 (1998).
79. Couch, W. J., Barger, A. J., Smail, I., Ellis, R. S. & Sharples, R. M. Morphological studies of the galaxy populations in distant 'Butcher-Oemler' clusters with the Hubble Space Telescope. II. AC 103, AC 118, and AC 114 at  $Z = 0.31$ . *Astrophys. J.* **497**, 188–211 (1998).
80. Postman, M., Lubin, L. M. & Oke, J. B. A study of nine high-redshift clusters of galaxies. II. Photometry, spectra, and ages of clusters 0023+0423 and 1604+4304. *Astron. J.* **116**, 560–583 (1998).
81. van Dokkum, P. G., Franx, M., Fabricant, D., Kelson, D. D. & Illingworth, G. D. A high merger fraction in the rich cluster MS 1054–03 at  $Z = 0.83$ : Direct evidence for hierarchical formation of massive galaxies. *Astrophys. J.* **520**, L95–L98 (1999).
82. Balogh, M. L. *et al.* Distinguishing local and global influences on galaxy morphology: A Hubble Space telescope comparison of high and low x-ray luminosity clusters. *Astrophys. J.* **566**, 123–136 (2002).
83. Lilly, S. *et al.* Hubble Space Telescope imaging of the CFRS and LDSS redshift surveys. II. Structural parameters and the evolution of disk galaxies to  $Z$  approximately 1. *Astrophys. J.* **500**, 75–94 (1998).
84. Simard, L. *et al.* The magnitude-size relation of galaxies out to  $z \sim 1$ . *Astrophys. J.* **519**, 563–579 (1999).
85. Lowenthal, J. D. *et al.* Keck spectroscopy of redshift  $Z$  approximately 3 Galaxies in the Hubble Deep Field. *Astrophys. J.* **481**, 673–688 (1997).
86. Steidel, C. C., Giallisco, M., Dickinson, M. & Adelberger, K. L. Spectroscopy of Lyman break galaxies in the Hubble Deep Field. *Astron. J.* **112**, 352–358 (1996).
87. Guzman, R. *et al.* The nature of compact galaxies in the Hubble Deep Field. II. Spectroscopic properties and implications for the evolution of the star formation rate density of the Universe. *Astrophys. J.* **489**, 559–572 (1997).
88. Connolly, A. J. *et al.* Slicing through multicolor space: Galaxy redshifts from broadband photometry. *Astron. J.* **110**, 2655–2664 (1995).
89. Hogg, D. W. *et al.* A blind test of photometric redshift prediction. *Astron. J.* **115**, 1418–1422 (1998).
90. Giallisco, M. *et al.* The Great Observatories Origins Deep Survey: Initial results from optical and near-infrared imaging. *Astrophys. J.* **600**, L93–L98 (2004).
91. Rix, H.-W. *et al.* GEMS: Galaxy evolution from morphologies and SEDs. *Astrophys. J. Suppl. Ser.* **152**, 163–173 (2004).
92. Davis, M. *et al.* The All-Wavelength Extended Groth Strip International Survey (AEGIS) data sets. *Astrophys. J.* **660**, L1–L6 (2007).
93. Scoville, N. *et al.* The Cosmic Evolution Survey (COSMOS): Overview. *Astrophys. J. Suppl. Ser.* **172**, 1–8 (2007).
94. Connolly, A. J., Szalay, A. S., Dickinson, M., Subbarao, M. U. & Brunner, R. J. The evolution of the global star formation history as measured from the Hubble Deep Field. *Astrophys. J.* **486**, L11–L14 (1997).
95. Sawicki, M. J., Lin, H. & Yee, H. K. C. Evolution of the Galaxy population based on photometric redshifts in the Hubble Deep Field. *Astron. J.* **113**, 1–12 (1997).
96. Fernández-Soto, A., Lanzetta, K. M. & Yahil, A. A new catalog of photometric redshifts in the Hubble Deep Field. *Astrophys. J.* **513**, 34–50 (1999).
97. Madau, P. *et al.* High-redshift galaxies in the Hubble Deep Field: colour selection and star formation history to  $z \sim 4$ . *Mon. Not. R. Astron. Soc.* **283**, 1388–1404 (1996).
98. Gallart, C., Zoccali, M. & Aparicio, A. The adequacy of stellar evolution models for the interpretation of the color-magnitude diagrams of resolved stellar populations. *Annu. Rev. Astron. Astrophys.* **43**, 387–434 (2005).
99. Holland, S., Fahlman, G. G. & Richer, H. B. Deep HST V- and I-Band observations of the halo of M31: Evidence for multiple stellar populations. *Astron. J.* **112**, 1035–1045 (1996).
100. Brown, T. M. *et al.* Evidence of a significant intermediate-age population in the M31 halo from main-sequence photometry. *Astrophys. J.* **592**, L17–L20 (2003).
101. Brown, T. M. *et al.* The detailed star formation history in the spheroid, outer disk, and tidal stream of the Andromeda galaxy. *Astrophys. J.* **652**, 323–353 (2006).
102. Brown, T. M. *et al.* The extended star formation history of the Andromeda spheroid at 21 kpc on the minor axis. *Astrophys. J.* **658**, L95–L98 (2007).
103. Harris, G. L. H., Harris, W. E. & Poole, G. B. The metallicity distribution in the halo stars of NGC 5128: Implications for galaxy formation. *Astron. J.* **117**, 855–867 (1999).
104. Harris, W. E. & Harris, G. L. H. The halo stars in NGC 5128. III. An inner halo field and the metallicity distribution. *Astron. J.* **123**, 3108–3123 (2002).
105. Barker, M. K., Sarajedini, A., Geisler, D., Harding, P. & Schommer, R. The stellar populations in the outer regions of M33. III. Star formation history. *Astron. J.* **133**, 1138–1160 (2007).



106. Williams, B. F. *et al.* The ACS Nearby Galaxy Survey Treasury I. The star formation history of the M81 outer disk. Preprint at (<http://arxiv1.library.cornell.edu/abs/0810.2557>) (2008).
107. Dohm-Palmer, R. C. *et al.* Deep Hubble Space Telescope imaging of Sextans A. I. The spatially resolved recent star formation history. *Astron. J.* **123**, 813–831 (2002).
108. Richer, H. B. *et al.* White dwarfs in globular clusters: Hubble Space Telescope observations of M4. *Astrophys. J.* **484**, 741–760 (1997).
109. Calamida, A. *et al.* On the white dwarf cooling sequence of the globular cluster  $\omega$  Centauri. *Astrophys. J.* **673**, L29–L33 (2008).
110. Hansen, B. M. S. *et al.* The white dwarf cooling sequence of the globular cluster Messier 4. *Astrophys. J.* **574**, L155–L158 (2002).
111. Cool, A. M., Piotto, G. & King, I. R. The main sequence and a white dwarf sequence in the globular cluster NGC 6397. *Astrophys. J.* **468**, 655–662 (1996).
112. Hansen, B. M. S. *et al.* Hubble Space Telescope observations of the white dwarf cooling sequence of M4. *Astrophys. J. Suppl. Ser.* **155**, 551–576 (2004).
113. Bedin, L. R. *et al.* The white dwarf cooling sequence in NGC 6791. *Astrophys. J.* **624**, L45–L48 (2005).
114. Hansen, B. M. S. *et al.* The white dwarf cooling sequence of NGC 6397. *Astrophys. J.* **671**, 380–401 (2007).
115. Bedin, L. R. *et al.* Reaching the end of the white dwarf cooling sequence in NGC 6791. *Astrophys. J.* **678**, 1279–1291 (2008).
116. Perlmutter, S. *et al.* Measurements of the cosmological parameters Omega and Lambda from the first seven supernovae at  $z \geq 0.35$ . *Astrophys. J.* **483**, 565–581 (1997).
117. Riess, A. G. *et al.* Observational evidence from supernovae for an accelerating universe and a cosmological constant. *Astron. J.* **116**, 1009–1038 (1998).
118. Garnavich, P. M. *et al.* Supernova limits on the cosmic equation of state. *Astrophys. J.* **509**, 74–79 (1998).
119. Perlmutter, S. *et al.* Measurements of Omega and Lambda from 42 high-redshift supernovae. *Astrophys. J.* **517**, 565–586 (1999).
120. Riess, A. G. *et al.* The farthest known supernova: Support for an accelerating universe and a glimpse of the epoch of deceleration. *Astrophys. J.* **560**, 49–71 (2001).
121. Knop, R. A. *et al.* New constraints on  $\Omega_M$ ,  $\Omega_\Lambda$  and  $w$  from an independent set of 11 high-redshift supernovae observed with the Hubble Space Telescope. *Astrophys. J.* **598**, 102–137 (2003).
122. Riess, A. G. *et al.* Type Ia supernova discoveries at  $z > 1$  from the Hubble Space Telescope: Evidence for past deceleration and constraints on dark energy evolution. *Astrophys. J.* **607**, 665–687 (2004).
123. Kneib, J.-P., Ellis, R. S., Smail, I., Couch, W. J. & Sharples, R. M. Hubble Space Telescope observations of the lensing cluster Abell 2218. *Astrophys. J.* **471**, 643–656 (1996).
124. Squires, G. *et al.* The dark matter, gas, and galaxy distributions in Abell 2218: A weak gravitational lensing and X-ray analysis. *Astrophys. J.* **461**, 572–586 (1996).
125. Hoekstra, H., Franx, M., Kuijken, K. & Squires, G. Weak lensing analysis of CL 1358+62 using Hubble Space Telescope observations. *Astrophys. J.* **504**, 636–660 (1998).
126. Hoekstra, H., Franx, M. & Kuijken, K. Hubble Space Telescope weak-lensing study of the  $z=0.83$  cluster MS 1054–03. *Astrophys. J.* **532**, 88–108 (2000).
127. Clowe, D. *et al.* A direct empirical proof of the existence of dark matter. *Astrophys. J.* **648**, L109–L113 (2006).
128. Bahcall, J. N. *et al.* The Hubble Space Telescope quasar absorption line key project. I – First observational results, including Lyman-alpha and Lyman-limit systems. *Astrophys. J. Suppl. Ser.* **87**, 1–43 (1993).
129. Morris, S. L., Weymann, R. J., Savage, B. D. & Gilliland, R. L. First results from the Goddard High-Resolution Spectrograph - The Galactic halo and the Ly-alpha forest at low redshift in 3C 273. *Astrophys. J.* **377**, L21–L24 (1991).
130. Penton, S. V., Stocke, J. T. & Shull, J. M. The local Ly $\alpha$  forest. IV. Space Telescope Imaging Spectrograph G140M spectra and results on the distribution and baryon content of H I absorbers. *Astrophys. J. Suppl. Ser.* **152**, 29–62 (2004).
131. Williger, G. M. *et al.* The low-redshift Ly $\alpha$  forest toward PKS 0405–123. *Astrophys. J.* **636**, 631–653 (2006).
132. Lehner, N. *et al.* Physical properties, baryon content, and evolution of the Ly $\alpha$  Forest: New insights from high-resolution observations at  $z \geq 0.4$ . *Astrophys. J.* **658**, 680–709 (2007).
133. Lanzetta, K. M., Bowen, D. V., Tytler, D. & Webb, J. K. The gaseous extent of galaxies and the origin of Lyman-alpha absorption systems: A survey of galaxies in the fields of Hubble Space Telescope spectroscopic target QSOs. *Astrophys. J.* **442**, 538–568 (1995).
134. Stocke, J. T., Shull, J. M., Penton, S., Donahue, M. & Carilli, C. The local Ly alpha forest: Association of clouds with superclusters and voids. *Astrophys. J.* **451**, 24–43 (1995).
135. Le Brun, V., Bergeron, J., Boisse, P. & Deharveng, J. M. The nature of intermediate-redshift damped Ly $\alpha$  absorbers. *Astron. Astrophys.* **321**, 733–748 (1997).
136. Steidel, C. C., Dickinson, M., Meyer, D. M., Adelberger, K. L. & Sembach, K. R. Quasar absorbing galaxies at  $z \geq 1$ . I. Deep imaging and spectroscopy in the field of 3C 336. *Astrophys. J.* **480**, 568–588 (1997).
137. Chen, H.-W., Lanzetta, K. M., Webb, J. K. & Barcons, X. The gaseous extent of galaxies and the origin of Ly alpha absorption systems. III. Hubble Space Telescope imaging of Ly alpha-absorbing galaxies at  $z < 1$ . *Astrophys. J.* **498**, 77–94 (1998).
138. Tripp, T. M., Lu, L. & Savage, B. D. The relationship between galaxies and low-redshift weak Ly alpha absorbers in the directions of H1821+643 and PG 1116+215. *Astrophys. J.* **508**, 200–231 (1998).
139. Tripp, T. M., Savage, B. D. & Jenkins, E. B. Intervening O VI quasar absorption systems at low redshift: A significant baryon reservoir. *Astrophys. J.* **534**, L1–L5 (2000).
140. Davé, R. *et al.* Baryons in the warm-hot intergalactic medium. *Astrophys. J.* **552**, 473–483 (2001).
141. Sembach, K. R. *et al.* Highly ionized high-velocity gas in the vicinity of the Galaxy. *Astrophys. J. Suppl. Ser.* **146**, 165–208 (2003).
142. Nicastro, F. *et al.* The far-ultraviolet signature of the ‘missing’ baryons in the Local Group of galaxies. *Nature* **421**, 719–721 (2003).
143. Savage, B. D. & Sembach, K. R. Interstellar abundances from absorption-line observations with the Hubble Space Telescope. *Annu. Rev. Astron. Astrophys.* **34**, 279–330 (1996).

**Acknowledgements** The author is happy to acknowledge many discussions with her colleagues, including H.-W. Rix, E. Bell, D. Hogg, A. Burgasser, G. Laughlin, S. Anderson, N. Reid, D. Schneider, D. Soderblom, S. Sigurdson, J. Rigby, P. Plait, M. Livio, J. Wisniewski, and the readers of Cosmic Variance. She also thanks the staff at the Max-Planck-Institut für Astronomie for their hospitality while this article was being written.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence should be addressed to J.D. (jd@astro.washington.edu).

# Cell-cycle restriction limits DNA damage and maintains self-renewal of leukaemia stem cells

Andrea Viale<sup>1\*</sup>, Francesca De Franco<sup>1,2\*</sup>, Annette Orleth<sup>1,2\*</sup>, Valeria Cambiaghi<sup>1</sup>, Virginia Giuliani<sup>1</sup>, Daniela Bossi<sup>1</sup>, Chiara Ronchini<sup>1</sup>, Simona Ronzoni<sup>1</sup>, Ivan Muradore<sup>1</sup>, Silvia Monestiroli<sup>1</sup>, Alberto Gobbi<sup>1</sup>, Myriam Alcalay<sup>1,4</sup>, Saverio Minucci<sup>1,3</sup> & Pier Giuseppe Pelicci<sup>1,4</sup>

**Rare cells with the properties of stem cells are integral to the development and perpetuation of leukaemias. A defining characteristic of stem cells is their capacity to self-renew, which is markedly extended in leukaemia stem cells. The underlying molecular mechanisms, however, are largely unknown. Here we demonstrate that expression of the cell-cycle inhibitor p21 is indispensable for maintaining self-renewal of leukaemia stem cells. Expression of leukaemia-associated oncogenes in mouse haematopoietic stem cells (HSCs) induces DNA damage and activates a p21-dependent cellular response, which leads to reversible cell-cycle arrest and DNA repair. Activated p21 is critical in preventing excess DNA-damage accumulation and functional exhaustion of leukaemic stem cells. These data unravel the oncogenic potential of p21 and suggest that inhibition of DNA repair mechanisms might function as potent strategy for the eradication of the slowly proliferating leukaemia stem cells.**

The number of times a single HSC replicates during a lifespan in mammals is limited (about 80–200), suggesting that self-renewal of HSCs is intrinsically restricted<sup>1–3</sup>. This limit becomes experimentally evident as exhaustion of their regenerative potential when HSCs are induced to proliferate rapidly in response to environmental stress, such as in serial transplantation or myelosuppressive chemotherapy<sup>4</sup>. Under these conditions, cell-cycle restriction of HSCs might become critical, because in mice selectively deficient for the cell-cycle inhibitor p21Cip1/Waf1 (p21<sup>-/-</sup>), HSCs proliferate more and are rapidly consumed<sup>5</sup>.

Like normal blood cells, leukaemias are maintained by rare stem cells (leukaemia stem cells (LSCs)), endowed with the capacity to self-renew and to generate many leukaemic blasts<sup>6</sup>. Self-renewal of LSCs is virtually unlimited, as inferred by their ability to support continuous expansion of the leukaemic clone and to be propagated inexhaustibly in mice during serial bone marrow transplantation<sup>7,8</sup>. It is not clear, however, whether specific mechanisms are selected during leukaemogenesis to protect hyper-proliferating LSCs from exhaustion. This paper investigates the role of p21Cip1/Waf1 (hereafter called p21) in the regulation of LSC self-renewal in mice, using PML–RAR, the initiating oncogene of human acute promyelocytic leukaemia (APL)<sup>9</sup>, one subtype of acute myeloid leukaemia (AML).

## p21 maintains self-renewal of PML–RAR HSCs

As model systems, we used mice expressing the PML–RAR complementary DNA from the endogenous cathepsin-G promoter ('knock-in' PML–RAR mice<sup>10</sup>), backcrossed in the wild-type (WT) or p21<sup>-/-</sup> C57BL6 backgrounds (PR and PRp21<sup>-/-</sup> mice). PR mice develop APL after a long latency and at a low frequency, suggesting that other genetic alterations cooperate with PML–RAR<sup>9,10</sup>. We analysed stem cells in the bone marrow of pre-leukaemic and leukaemic mice.

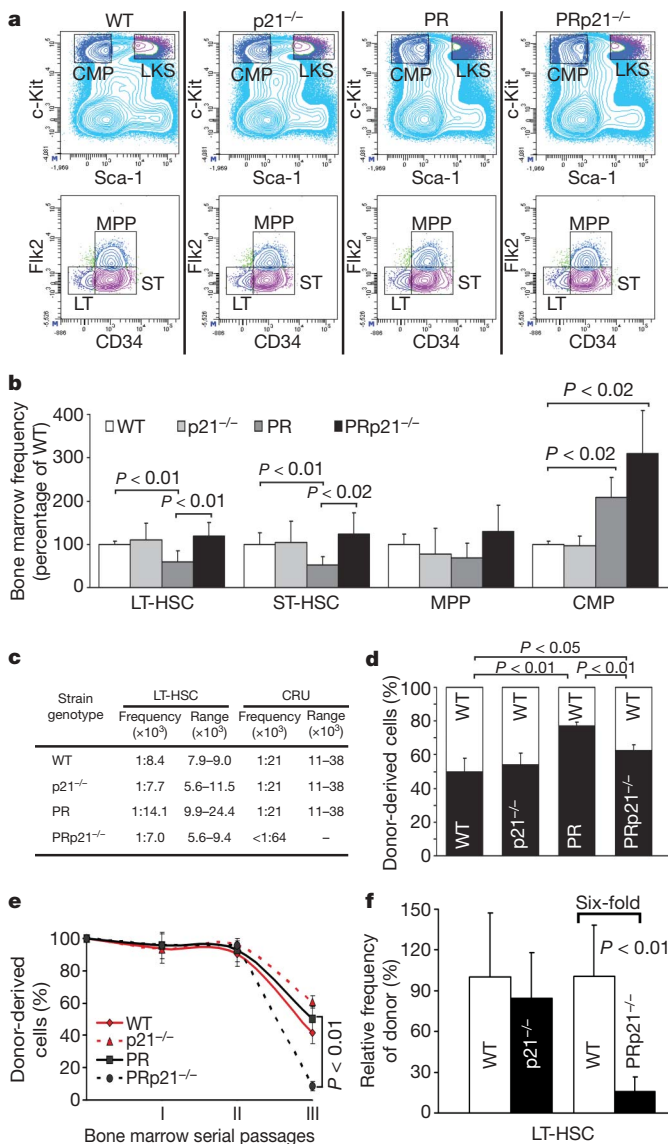
We first quantified the frequency of HSCs in the bone marrow of pre-leukaemic 2- to 3-month-old mice by immunophenotyping and limiting dilution. A unique cell surface phenotype (Lin<sup>-</sup>Kit<sup>+</sup>Sca1<sup>+</sup>Flk2<sup>-</sup>CD34<sup>-</sup>) identifies a sub-population that is highly enriched with cells capable of long-term reconstitution of the haematopoietic system<sup>11–13</sup> (long-term reconstituting HSCs (LT-HSCs); see Fig. 1a and Supplementary Fig. 2 for representative fluorescence-activated cell sorting (FACS) profiles and expression of PML–RAR in HSCs). For limiting dilutions<sup>14,15</sup>, we performed competitive transplantation of three concentrations of test bone marrow cells (500 × 10<sup>3</sup>, 100 × 10<sup>3</sup> and 10 × 10<sup>3</sup>) using the CD45 (Ly5.1 or Ly5.2) congenic system, and estimated the frequency of competitive re-populating units (CRUs) by Poisson statistics. The calculated frequencies of WT LT-HSCs (Fig. 1b, c) and CRUs (Fig. 1c and Supplementary Fig. 3) were, respectively, 1:8.4 × 10<sup>3</sup> and 1:21 × 10<sup>3</sup>, as reported<sup>14,16–19</sup>. Although in the p21<sup>-/-</sup> bone marrow we observed comparable frequencies, in the PR bone marrow the frequency of LT-HSCs was slightly diminished (1:14.1 × 10<sup>3</sup>; Fig. 1b, c), although CRUs were unaffected (1:21 × 10<sup>3</sup>; Fig. 1c and Supplementary Fig. 3). This apparent discrepancy might reflect different sensitivities of the two assays. Surprisingly, in the PRp21<sup>-/-</sup> bone marrow we found normal frequencies of LT-HSCs (1:7.0 × 10<sup>3</sup>) and a marked decrease of CRUs (less than 1:64 × 10<sup>3</sup>; Fig. 1c and Supplementary Fig. 3). Thus, p21 loss or PML–RAR expression does not significantly affect the establishment or maintenance of HSC reserves, whereas p21 expression is critical in maintaining numbers of functional HSCs in the presence of PML–RAR.

To investigate the functional properties of PRp21<sup>-/-</sup> HSCs, we evaluated their capacity to re-populate an irradiated host by competitive bone marrow transplantation. Quantification of bone marrow reconstitution of mice inoculated with about one PRp21<sup>-/-</sup> CRU

<sup>1</sup>Department of Experimental Oncology at the IFOM-IEO Campus, European Institute of Oncology, IEO, 20141 Milan, Italy. <sup>2</sup>Dipartimento di Medicina Clinica e Sperimentale, Policlinico Monteluce, University of Perugia, 06100 Perugia, Italy. <sup>3</sup>Dipartimento di Scienze Biomolecolari e Biotecnologie, University of Milano 20100, Milano, Italy. <sup>4</sup>Dipartimento di Medicina, Chirurgia e Odontoiatria, University of Milano, Milano 20100, Italy.

\*These authors contributed equally to this work.





**Figure 1 | p21 maintains self-renewal of PML-RAR-expressing HSCs.**

**a, b**, LT-HSCs (LT), ST-HSCs (ST), multipotent progenitors (MPP) and CMP: representative profiles (**a**) and frequencies (**b**) of indicated genotypes (mean and s.d. of five experiments; four mice per group). **c**, Frequency of LT-HSCs and CRUs. **d, e**, Frequency of peripheral blood or bone marrow donor-derived cells after bone marrow competition (**d**) or serial (**e**) transplantation, respectively. Like others<sup>45</sup>, we did not observe premature exhaustion of p21<sup>-/-</sup> HSCs using the C57BL/6 strain, as instead reported for the 129Sv<sup>+</sup>. **f**, Frequency of donor bone marrow LT-HSCs after 8 weeks transplantation using equivalent LT-HSC numbers (mean and s.d. of two experiments; three mice per genotype).

(100 × 10<sup>3</sup> test bone marrow cells in the limiting dilution assay described above) revealed a slightly higher number of re-populating units (PRp21<sup>-/-</sup> to WT RU ratio = 1.7). Similarly, the re-populating unit ratio in a 1:1 admixture (1 × 10<sup>6</sup> cells) of PRp21<sup>-/-</sup> and competing WT bone marrow cells was 1.6 ( $P < 0.05$ ), with 62.5% of PRp21<sup>-/-</sup> cells in the peripheral blood of long-term reconstituted animals (Fig. 1d and Supplementary Fig. 4a). Thus, although numbers of functional HSCs are reduced in the PRp21<sup>-/-</sup> bone marrow, they possess normal, or slightly increased, re-populating capacity.

To investigate alternative mechanisms for the reduced pool of functional HSCs in the PRp21<sup>-/-</sup> bone marrow, we analysed their lifespan by serial transplantation. Briefly, test Ly5.2<sup>+</sup> bone marrow cells were injected into irradiated Ly5.1<sup>+</sup> mice, analysed 3–4 months post-transplant and re-transplanted. Bone marrow reconstitution with WT, PR or p21<sup>-/-</sup> donor cells was about 50% after the third transplant

and, strikingly, only about 8% with the PRp21<sup>-/-</sup> bone marrow (Fig. 1e), indicating premature exhaustion of the PRp21<sup>-/-</sup> HSCs.

To determine whether the functional decline of PRp21<sup>-/-</sup> HSCs was due to diminished self-renewal capacity, we assayed their proliferative potential in primary transplant recipients and *in vitro*. Equal numbers of test Ly5.2<sup>+</sup> LT-HSCs were injected into irradiated Ly5.1<sup>+</sup> mice and bone marrows analysed immunophenotypically after two months. The numbers of donor-derived PRp21<sup>-/-</sup> LT-HSCs was about sixfold less, compared with PR LT-HSCs ( $P < 0.01$ ; Fig. 1f). Likewise, the total progeny-cell output of PRp21<sup>-/-</sup> HSCs after short-term culture was significantly reduced (about fourfold; Supplementary Fig. 5), demonstrating that PRp21<sup>-/-</sup> HSCs have diminished self-renewal capacity. Notably, p21 loss had no significant effects on the proliferation of LT-HSCs in the absence of PML-RAR expression, either *in vivo* (Fig. 1f) or *in vitro* (Supplementary Fig. 5).

Cumulatively, these results suggest that p21 is critical to the self-renewal of PML-RAR-expressing HSCs and to protection from functional exhaustion in stressful conditions. Because PML-RAR expression or p21 loss alone did not reduce the self-renewal capacity of HSCs, their synergistic effect suggests that they affect a single function or pathway.

### p21 limits DNA-damage of PML-RAR HSCs

Functional exhaustion of HSCs was recently described in mice deficient in several genomic-maintenance pathways, owing to accumulation of genomic damage<sup>20,21</sup>. Thus, we examined whether PML-RAR expression and/or p21 loss induce accumulation of DNA damage by immunostaining for phosphorylation of histone H2AX ( $\gamma$ H2AX), as an indicator of the cellular response to DNA damage<sup>22</sup>. Expression of PML-RAR in U937 haematopoietic cells induced marked accumulation of  $\gamma$ H2AX foci (Fig. 2a), suggesting that the fusion protein induces DNA damage. The same effect was observed in the HSC-enriched populations of murine (lineage depleted: Lin<sup>-</sup>; Fig. 2b) or human (CD34<sup>+</sup>; Fig. 2c) origin.

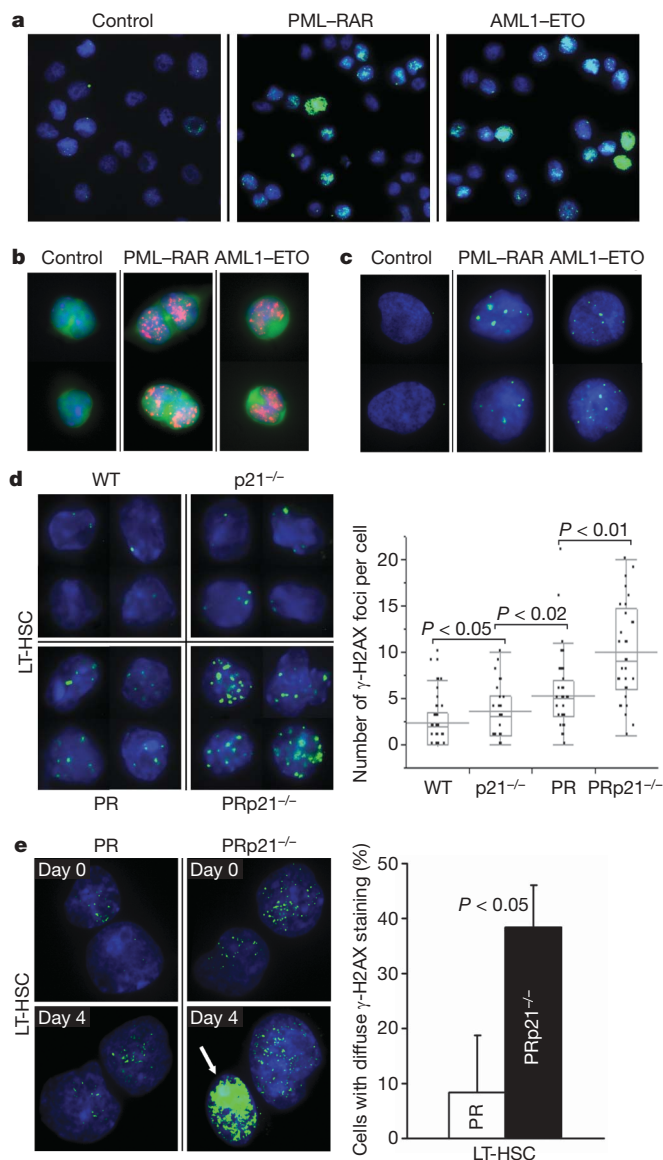
We then investigated the effect of PML-RAR and p21 on DNA integrity *in vivo*, by quantifying  $\gamma$ H2AX foci in the LT-HSCs of 2- to 3-month-old mice (Fig. 2d and Supplementary Fig. 6). Whereas rare  $\gamma$ H2AX foci were observed in WT LT-HSCs (about 70% of cells with no more than two foci per cell; modal number = 0), as reported<sup>20</sup>, most (about 70%) of the PR LT-HSCs presented multiple foci (at least two; modal number = 5). p21 loss also led to increased LT-HSCs with  $\gamma$ H2AX foci, although at a lower frequency (about 60% of cells with at least two foci per cell; modal number = 1). Strikingly, PRp21<sup>-/-</sup> LT-HSCs were almost homogeneously positive for anti- $\gamma$ H2AX staining (more than 95% of cells with at least two foci per cell; modal number = 7). Together, these results demonstrate that either PML-RAR expression or p21 loss induces a moderate degree of DNA-damage in LT-HSCs, which increases significantly when both conditions are present, suggesting that p21 limits DNA-damage accumulation in PML-RAR-expressing HSCs.

The synthetic interaction between p21 loss and PML-RAR expression on HSC functions becomes evident when HSCs are induced to proliferate, as it occurs after bone marrow transplantation. Because hyperproliferation causes DNA-replication stress and genomic damage<sup>23</sup>, we investigated DNA-damage accumulation in actively proliferating HSCs. Analysis of  $\gamma$ H2AX foci in cultured LT-HSCs showed increased time-dependent DNA-damage in both PR and PRp21<sup>-/-</sup> samples, which was significantly higher in the latter (Fig. 2e). Thus, p21 limits accumulation of DNA damage during proliferation of PML-RAR-expressing HSCs.

### p21 is upregulated by PML-RAR and induces cell-cycle restriction of HSCs

p21 is upregulated by DNA damage or other types of stress and induces reversible cell-cycle arrest, allowing for DNA repair or stress resistance<sup>24–26</sup>. We thus investigated the effects of PML-RAR on p21 expression and its consequences on the cell cycle.

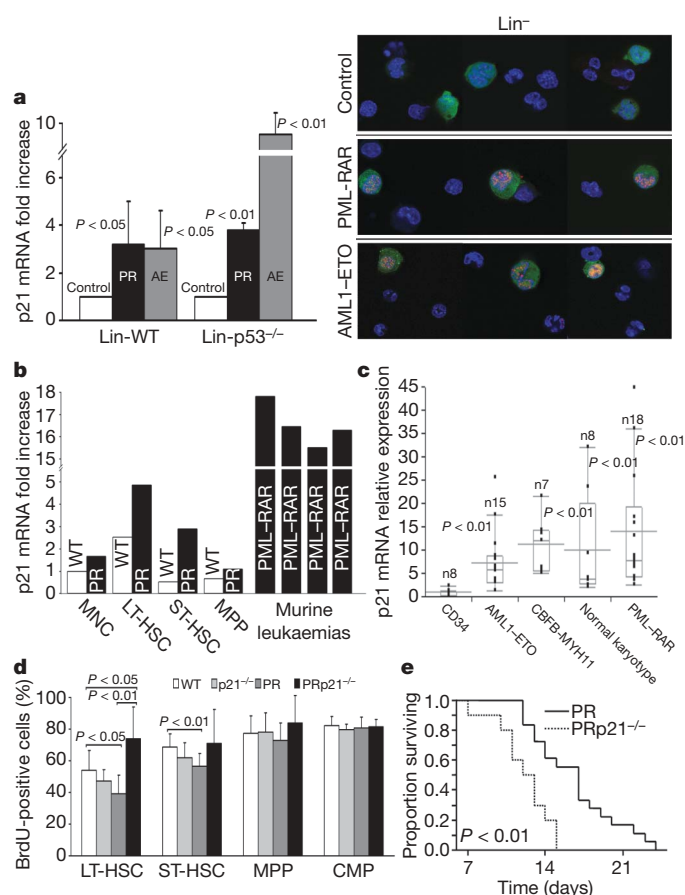
Protein and/or RNA analyses showed increased levels of p21 in U937 (Supplementary Fig. 7a–c) and Lin<sup>-</sup> (Fig. 3a) cells after PML-RAR expression, as well as in the LT-HSCs of the PR bone marrow



**Figure 2 | Fusion proteins induce DNA damage.** Anti- $\gamma$ H2AX immunofluorescence: **a**, different U937 clones ( $\alpha$ - $\gamma$ H2AX: green; DAPI: blue;  $\times 40$ ); **b**, Lin<sup>-</sup> cells infected with empty (control), PML-RAR or AML1-ETO-expressing green fluorescent protein (GFP) retroviruses (GFP: green;  $\alpha$ - $\gamma$ H2AX: red; DAPI: blue;  $\times 125$ ); **c**, CD34<sup>+</sup> cells infected with empty (control), PML-RAR or AML1-ETO retroviruses ( $\alpha$ - $\gamma$ H2AX: green; DAPI: blue;  $\times 125$ ); **d**, LT-HSCs of indicated genotypes ( $\alpha$ - $\gamma$ H2AX: green; DAPI: blue;  $\times 125$ ; left), numbers of  $\gamma$ H2AX foci per cell (right) (error bars denote s.d.); **e**, PR or PRp21<sup>-/-</sup> LT-HSCs before (day 0) or after 4 days of culture ( $\alpha$ - $\gamma$ H2AX: green; DAPI: blue;  $\times 125$ ; left). Frequencies of cells showing diffuse staining (arrow marks one representative cell) (right). Mean and s.d. of three experiments are shown.

(Fig. 3b), demonstrating that PML-RAR upregulates p21. Notably, we also observed increased levels of p21 expression in freshly isolated blasts from murine (Fig. 3b) or human (Fig. 3c) APLs, indicating that the effect of PML-RAR on p21 is retained in the fully transformed APL blasts. This effect was surprising, because PML-RAR downregulates p53, a critical activator of p21 expression<sup>27</sup>. Analysis of the effects of PML-RAR on p21 in p53<sup>-/-</sup> Lin<sup>-</sup> showed that, indeed, p21 upregulation by PML-RAR is p53 independent (Fig. 3a).

The effects of p21 on the cell cycle of HSCs were investigated *in vivo* using the 5-bromodeoxyuridine (BrdU) incorporation test. The fraction of BrdU-positive LT-HSCs was decreased in the PR bone marrow ( $P < 0.05$ ), whereas it was increased in the PRp21<sup>-/-</sup> bone marrow ( $P < 0.01$ ) (Fig. 3d and Supplementary Fig. 8a), demonstrating that p21 induces cell-cycle restriction of PML-RAR-expressing LT-HSCs. This



**Figure 3 | p21 upregulation and cell-cycle restriction in HSCs.**

**a**, Quantitative PCR (qPCR) (left; mean and s.d. of three experiments) and confocal microscopy (right) of p21 expression in Lin<sup>-</sup> cells infected with empty (control), PML-RAR (PR) or AML1-ETO (AE) GFP-retroviruses ( $\alpha$ -p21: red; DAPI: blue; GFP: green; pixel width: 103–132 nm). **b**, **c**, qPCR of p21 expression in MNCs, bone marrow sub-populations and mouse PML-RAR leukaemias (relative to WT MNCs) (**b**) and in healthy CD34<sup>+</sup> cells and AML-patient blasts (expressing AML1-ETO, CBF $\beta$ -MYH11, PML-RAR or with normal karyotypes) (**c**). **d**, BrdU<sup>+</sup> cells in bone marrow sub-populations (mean and s.d. of three experiments; three or four mice per genotype). **e**, Survival curves of PR ( $n = 18$ ) and PRp21<sup>-/-</sup> ( $n = 11$ ) mice during 5-FU treatment.

effect of p21 correlates with the observed reduction in LT-HSC frequency in the PR bone marrow (Fig. 1b, c). For functional confirmation, we measured the sensitivity of PR and PRp21<sup>-/-</sup> mice to myelosuppression, using a cell-cycle-specific drug (5-fluorouracil, 5-FU). The survival outcome of PRp21<sup>-/-</sup> mice was significantly lower than that of the PR mice (Fig. 3e), thus confirming that the frequency of cell-cycle-restricted HSCs is lower in the PRp21<sup>-/-</sup> bone marrow.

Together, these data demonstrate that expression of PML-RAR in HSCs results in upregulation of p21, which, in turn, induces cell-cycle restriction. They also suggest that this function of p21 limits DNA damage accumulation by favouring DNA repair. It remains unclear, however, whether the effect of PML-RAR on p21 is direct, the consequence of PML-RAR-induced DNA damage, or both.

### p21 prevents hyper-proliferation, severe DNA-damage and exhaustion of LSCs

We then investigated whether p21 expression and cell-cycle restriction are critical in maintaining self-renewal of LSCs. PR and PRp21<sup>-/-</sup> mice developed APL with comparable latency, frequency and morphology (Fig. 4a, b), indicating that p21 is dispensable for leukaemogenesis by PML-RAR. Strikingly, although the PR and PRp21<sup>-/-</sup> blasts showed comparable homing properties (Supplementary Fig. 9), none of the PRp21<sup>-/-</sup> leukaemias was able to propagate the disease when transplanted into syngeneic recipients (Fig. 4c). To confirm these results in



another model system of APL<sup>9</sup>, WT and p21<sup>-/-</sup> Lin<sup>-</sup> cells were infected with a PML-RAR-encoding retrovirus and transplanted into recipient mice. Both groups of mice reconstituted haematopoiesis and developed APL with comparable latency and frequency (Supplementary Fig. 10a–c). Again, however, no p21<sup>-/-</sup> APL was transplantable (Supplementary Fig. 10d), confirming that p21 is required for PML-RAR LSCs to support leukaemia outgrowth after transplantation.

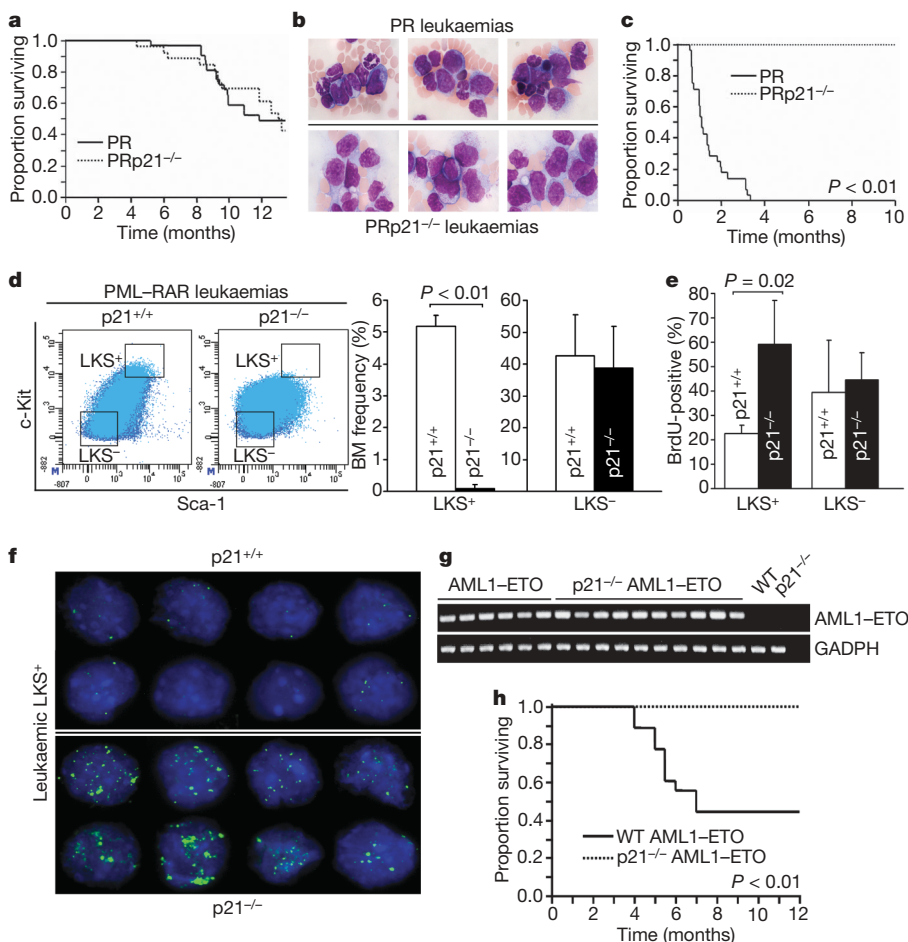
We next examined the effects of p21 on numbers, cell-cycle properties and DNA damage of LSCs, by analysing a sub-population of Lin<sup>-</sup> blasts that expresses the c-Kit and Sca1 HSC-markers (LKS<sup>+</sup>; Fig. 4d) and is enriched with cells capable of propagating leukaemias upon transplantation (unpublished observations). Remarkably, PRp21<sup>-/-</sup> LKS<sup>+</sup> cells were reduced in number (about 50-fold; Fig. 4d), showed increased BrdU incorporation (about threefold; Fig. 4e and Supplementary Fig. 8b) and increased DNA damage (Fig. 4f). Thus, like in pre-leukaemic HSCs, p21 expression in LSCs imposes cell-cycle restriction, limits accumulation of DNA damage and prevents their exhaustion.

### p21 response to PML-RAR expression is specific to stem cells

Enforced expression of activated oncogenes in primary cells, such as fibroblasts, induces DNA damage and triggers a DNA-damage response that leads to a halt in proliferation, owing to apoptosis or irreversible cell-cycle arrest (senescence), thus preventing propagation of DNA alterations<sup>28–30</sup>. This cellular response depends on p53 and is considered a powerful barrier to tumour development<sup>31,32</sup>. Like other oncogenes, expression of PML-RAR induces DNA damage in HSCs, without, however, leading to exclusion of damaged cells. Indeed, DNA is damaged in PML-RAR-expressing HSCs (Fig. 2d), yet they maintain intact self-renewal (Fig. 1f). We investigated whether this response to PML-RAR is specific to HSCs, by analysing its effects on downstream progenitors and fibroblasts.

Expression of PML-RAR in common myeloid progenitors (CMPs) induced accumulation of  $\gamma$ H2AX foci (Fig. 5a) and upregulation of p21 (data not shown). Notably, the effect of the fusion protein on DNA integrity was independent of p21 (Fig. 5a). Immunophenotypic and BrdU-incorporation analyses showed, in both PR and PRp21<sup>-/-</sup> bone marrow, increased CMP numbers (Fig. 1b) in the absence of any significant effect on their cycling properties (Fig. 3d). This effect of PML-RAR on CMP numbers was confirmed by the colony-forming cell (CFC) assay (Fig. 5b) and might explain the better performance of PR bone marrow after competitive transplantation (Fig. 1d). Because PML-RAR inhibits apoptosis of haematopoietic progenitors by inhibiting p53 (ref. 27), the expansion of CMPs in PR bone marrow might be due to their prolonged survival. Thus, upregulation of p21 in myeloid progenitors has no effect on their cycling properties, nor does it regulate DNA-damage accumulation. Likewise, we observed no effects of p21 expression on the frequency (Fig. 4d), cell-cycle properties (Fig. 4e) and extent of DNA damage (Fig. 5c) of sub-populations of leukaemia blasts that do not express stem-cell markers (LSK<sup>-</sup>).

Expression of PML-RAR in mouse embryonic fibroblasts (MEFs) and human (WI-38) fibroblasts also induced accumulation of  $\gamma$ H2AX foci and upregulation of p21 (Fig. 5d and data not shown). In both cell types, PML-RAR blocked proliferation and induced  $\beta$ -galactosidase positivity (Fig. 5e–g and data not shown), a marker of cellular senescence<sup>33</sup>. This effect of the fusion protein was p21 dependent and p53 independent, as shown by the opposite behaviours of p21<sup>-/-</sup> or p53<sup>-/-</sup> MEFs after PML-RAR expression (Fig. 5e, f). These data demonstrate that although PML-RAR induces DNA damage and p21 upregulation in all tested target cells, p21 mediates different cellular responses in HSCs (cell-cycle restriction) and fibroblasts (permanent cell-cycle arrest). Thus, the p21 property

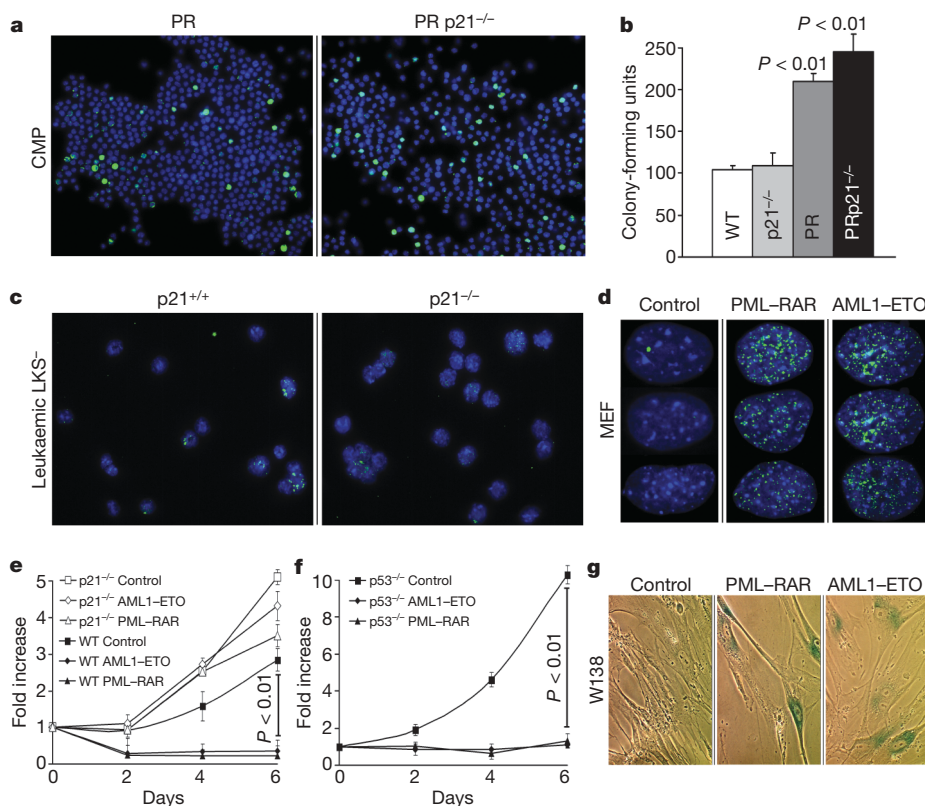


**Figure 4 | p21 limits DNA damage in LSCs and is indispensable for leukaemogenesis.** **a**, Survival of PR ( $n = 33$ ) and PRp21<sup>-/-</sup> ( $n = 26$ ) mice.

**b**, MGG blood smears of leukaemic mice.

**c**, Survival of secondary recipients (four each) injected with PR ( $n = 7$ ) or PRp21<sup>-/-</sup> ( $n = 7$ ) leukaemias. **d**, Leukaemic LKS<sup>+</sup> and LKS<sup>-</sup> subpopulations: representative profiles (left) and frequencies (right) of indicated samples (mean of four genotypes and s.d.). **e**, *In vivo* BrdU incorporation in LKS<sup>+</sup> and LKS<sup>-</sup> subpopulations (mean of three leukaemias and s.d.).

**f**, Anti- $\gamma$ H2AX staining of LKS<sup>+</sup> from PML-RAR leukaemias ( $\alpha$ - $\gamma$ H2AX: green; DAPI: blue;  $\times 125$ ). **g**, PCR of PB DNA for retrovirus integrations (**g**) and survival (**h**) of mice injected with WT ( $n = 18$ ) or p21<sup>-/-</sup> ( $n = 17$ ) AML1-ETO-infected Lin<sup>-</sup> cells.



**Figure 5 | Effects of fusion proteins on progenitors and fibroblasts.** Anti- $\gamma$ H2AX staining of **(a)** PR and PRp21<sup>-/-</sup> CMP ( $\times 10$ ) and **(c)** PML-RAR leukaemic LKS<sup>-</sup> ( $\times 30$ ) ( $\alpha$ - $\gamma$ H2AX: green; DAPI: blue); **b**, CFUs from bone marrow MNCs (mean  $\pm$  s.d. of three experiments). **d**, Anti- $\gamma$ H2AX staining of MEFs ( $\alpha$ - $\gamma$ H2AX: green; DAPI: blue;  $\times 125$ ). **e**, Growth curves of WT (solid symbols) and p21<sup>-/-</sup> (open symbols) MEFs expressing empty vector (control), PML-RAR or AML1-ETO (mean  $\pm$  s.d. of three experiments). **f**, Same as in **e** for p53<sup>-/-</sup> MEF (mean  $\pm$  s.d. of three experiments). **g**,  $\beta$ -Galactosidase staining of WI38 cells expressing fusion proteins.

of allowing expansion of DNA-damaged cells in response to PML-RAR expression is specific to HSCs. It remains unclear, however, whether this reflects a physiological response of HSCs to DNA damage or a selected property of the PML-RAR oncogene in HSCs.

### p21 is indispensable for leukaemogenesis by AML1-ETO

AML1-ETO, another AML-specific oncogene<sup>34</sup>, induced accumulation of  $\gamma$ H2AX foci and upregulation of p21 in U937 (Fig. 2a and Supplementary Fig. 7a–c) and Lin<sup>-</sup> (Figs 2b and 3a) cells, and in MEFs (Fig. 5d). Upregulation of p21 in Lin<sup>-</sup> cells was p53 independent (Fig. 3a). As for PML-RAR, AML1-ETO induced cell-cycle restriction of U937 cells (Supplementary Fig. 7d and data not shown) and senescence of cultured MEFs, in a p21-dependent and p53-independent manner (Fig. 5e, f), although it did not arrest proliferation of stem cell/progenitors, as revealed, for example, by the long-term culture-initiating cell (LTC-IC) assay (Supplementary Fig. 4b). Strikingly, AML1-ETO failed to induce leukaemias when expressed in p21<sup>-/-</sup> Lin<sup>-</sup> cells. Lin<sup>-</sup> cells from WT or p21<sup>-/-</sup> mice were infected with AML1-ETO-expressing retroviruses and used to reconstitute recipient mice. Analysis of AML1-ETO integration in the peripheral blood DNA of the reconstituted mice demonstrated engraftment in all transplanted mice (Fig. 4g). However, 10 of 18 mice (about 55%) reconstituted with AML1-ETO-expressing WT Lin<sup>-</sup> cells developed AML with a latency of 4–7 months, as expected (Fig. 4h and Supplementary Fig. 10b), whereas none of the 17 mice reconstituted with AML1-ETO-expressing p21<sup>-/-</sup> Lin<sup>-</sup> developed leukaemia over a period of 12–14 months (Fig. 4h and Supplementary Fig. 10b), thus indicating that p21 is critical to the property of AML1-ETO in initiating leukaemogenesis. The different effects of p21 loss on PML-RAR and AML1-ETO leukaemogenesis (exhaustion of established leukaemias or inhibition of initiation, respectively) might reflect different effects on the lifespans of the corresponding LSCs. Notably, the protocol of AML1-ETO-induced leukaemogenesis involves treatment of recipient mice with a potent DNA-damaging agent (nitrosourea)<sup>35,36</sup>, which might result in accelerated exhaustion of the AML1-ETO LSCs in the absence of p21.

### Conclusions

Our data provide the first demonstration that expression of a cell-cycle inhibitor, such as p21, is critical for the initiation and maintenance of leukaemogenesis. They are consistent with a model whereby expression of leukaemia-associated oncogenes in HSCs induces DNA damage and activates a p21-dependent cellular response that, in turn, imposes cell-cycle restriction and triggers repair of the damaged DNA (Supplementary Fig. 1). This effect of p21 has two consequences, both of which might be relevant for leukaemogenesis. The first is the accumulation of genomically unstable HSCs. In the presence of oncogenic fusion proteins, HSCs accumulate moderate degrees of DNA damage, which represents the balance of the DNA-damaging effects of the fusion proteins, and, seemingly, the DNA-repair activity of upregulated p21. Both of these conditions are known to accelerate the rate of mutagenesis<sup>37,38</sup>, thus providing an explanation for the role of p21 in initiating leukaemogenesis. The second consequence is a bypass of the physiological exhaustion of HSC self-renewal, which occurs in time owing to accumulation of DNA damage<sup>5,20,21</sup>. This effect of p21 confers an advantage to HSCs when they hyper-proliferate, as it occurs during stress or after full transformation (for example, in the LSCs), thus explaining the role of p21 in the maintenance of leukaemias (Supplementary Fig. 1). In a way, these activities of p21 represent a novel mechanistic contribution to tumorigenesis: overexpression of p21, in fact, is not transforming per se; it is, however, 'permissive' for leukaemia to progress.

These data bear important implications for the treatment of leukaemias. First, they imply that cell-cycle-restricted LSCs are critical for the initiation and/or maintenance of the leukaemic clone, suggesting that targeting this compartment might be critical to disease eradication. Although this concept was anticipated by the demonstration that human LSCs slowly proliferate<sup>39,40</sup> (see also Supplementary Fig. 11a–c for the mouse LSCs), our data provide one underlying molecular mechanism, thus paving the way to novel anti-leukaemic strategies. Second, they suggest that inhibition of DNA repair might be synthetic lethal with oncogene expression. A range of selective inhibitors of different DNA damage response pathways have now been identified, which might become part of new approaches to targeted cancer therapy.



## METHODS SUMMARY

Fusion proteins were expressed into Lin<sup>−</sup> and CD34<sup>+</sup> cells using the PINCO retroviral vector<sup>41,42</sup>. Expression of PML-RAR or AML1-ETO in PR9 or U937-AE cells was obtained using 100 μM ZnSO<sub>4</sub> (ref. 43). WT MEFs or WI38 cells were infected with PML-RAR or AML1-ETO-expressing pBabe retroviruses and puromycin-selected for 4 days. Leukaemic mice were generated by retroviral transduction of PML-RAR or AML1-ETO into Lin<sup>−</sup> bone marrow cells<sup>9,44</sup> of 129Sv mice. PML-RAR knock-in mice<sup>10</sup> were provided by T. J. Ley and backcrossed in the C57BL/6J background (PR). Mating PR and p21<sup>−/−</sup> C57BL/6J mice generated PRp21<sup>−/−</sup> mice. The various stem cell/progenitor sub-populations (LT-HSC, short-term HSC (ST-HSC), multipotent and common myeloid progenitors (MPP and CMP)) were identified by FACS or isolated by cell-sorting after staining of mouse bone marrow mononuclear cells (MNCs) with fluorochrome-conjugated antibodies against Sca1, c-Kit, Flk-2, CD34 or the lineage markers Cd11b, GRI, Ter-119, IL7-R, CD3, CD4, CD8 and B220. Frequency of functional LT-HSCs was determined by limiting bone marrow transplantation, injecting decreasing numbers of Ly5.2<sup>+</sup> test bone marrow MNCs and a fixed number of WT Ly5.1<sup>+</sup> bone marrow MNCs into Ly5.1-irradiated mice. Lifespan of HSCs was determined by serial bone marrow transplantation, injecting Ly5.2<sup>+</sup> test bone marrow MNCs into Ly5.1<sup>+</sup> recipients, and re-transplanting bone marrow MNCs of the reconstituted mice. Self-renewal of HSCs was determined by an equivalent transplantation assay, measuring the frequency of Ly5.2<sup>+</sup> donor-derived LT-HSCs in Ly5.1<sup>+</sup> recipients, transplanted with stem-cell equivalents of test and control bone marrow-MNCs. HSC re-populating potential was determined by a bone marrow competition assay, transplanting equal numbers of test Ly5.2<sup>+</sup> and competitor Ly5.1<sup>+</sup> HSCs. Cell-cycle properties of HSCs, LSCs or progenitor sub-populations were determined by the *in vivo* BrdU-incorporation test. Cell-cycles of HSCs were analysed *in vivo* by the 5-FU test, measuring survival of animals treated with weekly doses of the drug.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 2 August; accepted 29 October 2008.**

- Harrison, D. E. Long-term erythropoietic repopulating ability of old, young, and fetal stem cells. *J. Exp. Med.* **157**, 1496–1504 (1983).
- Harrison, D. E. & Astle, C. M. Loss of stem cell repopulating ability upon transplantation. Effects of donor age, cell number, and transplantation procedure. *J. Exp. Med.* **156**, 1767–1779 (1982).
- Shepherd, B. E. *et al.* Hematopoietic stem cell behavior in non-human primates. *Blood* **110**, 1806–1813 (2007).
- Cheng, T. Cell cycle inhibitors in normal and tumor stem cells. *Oncogene* **23**, 7256–7266 (2004).
- Cheng, T. *et al.* Hematopoietic stem cell quiescence maintained by p21cip1/waf1. *Science* **287**, 1804–1808 (2000).
- Bonnet, D. & Dick, J. E. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nature Med.* **3**, 730–737 (1997).
- Hope, K. J., Jin, L. & Dick, J. E. Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity. *Nature Immunol.* **5**, 738–743 (2004).
- Warner, J. K., Wang, J. C., Hope, K. J., Jin, L. & Dick, J. E. Concepts of human leukemic development. *Oncogene* **23**, 7164–7177 (2004).
- Minucci, S. *et al.* PML-RAR induces promyelocytic leukemias with high efficiency following retroviral gene transfer into purified murine hematopoietic progenitors. *Blood* **100**, 2989–2995 (2002).
- Westervelt, P. *et al.* High-penetrance mouse model of acute promyelocytic leukemia with very low levels of PML-RAR $\alpha$  expression. *Blood* **102**, 1857–1865 (2003).
- Christensen, J. L. & Weissman, I. L. Flk-2 is a marker in hematopoietic stem cell differentiation: a simple method to isolate long-term stem cells. *Proc. Natl Acad. Sci. USA* **98**, 14541–14546 (2001).
- Osawa, M., Hanada, K., Hamada, H. & Nakauchi, H. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science* **273**, 242–245 (1996).
- Yang, L. *et al.* Identification of Lin<sup>−</sup>Sca1<sup>+</sup>Kit<sup>+</sup>CD34<sup>+</sup>Flt3<sup>−</sup> short-term hematopoietic stem cells capable of rapidly reconstituting and rescuing myeloablated transplant recipients. *Blood* **105**, 2717–2723 (2005).
- Szilvassy, S. J., Humphries, R. K., Lansdorp, P. M., Eaves, A. C. & Eaves, C. J. Quantitative assay for totipotent reconstituting hematopoietic stem cells by a competitive repopulation strategy. *Proc. Natl Acad. Sci. USA* **87**, 8736–8740 (1990).
- Szilvassy, S. J., Lansdorp, P. M., Humphries, R. K., Eaves, A. C. & Eaves, C. J. Isolation in a single step of a highly enriched murine hematopoietic stem cell population with competitive long-term repopulating ability. *Blood* **74**, 930–939 (1989).
- Janzen, V. *et al.* Stem-cell ageing modified by the cyclin-dependent kinase inhibitor p16INK4a. *Nature* **443**, 421–426 (2006).
- Passegue, E., Wagers, A. J., Giuriato, S., Anderson, W. C. & Weissman, I. L. Global analysis of proliferation and cell cycle gene expression in the regulation of hematopoietic stem and progenitor cell fates. *J. Exp. Med.* **202**, 1599–1611 (2005).
- Rossi, D. J. *et al.* Cell intrinsic alterations underlie hematopoietic stem cell aging. *Proc. Natl Acad. Sci. USA* **102**, 9194–9199 (2005).
- Sudo, K., Ema, H., Morita, Y. & Nakauchi, H. Age-associated characteristics of murine hematopoietic stem cells. *J. Exp. Med.* **192**, 1273–1280 (2000).
- Rossi, D. J. *et al.* Deficiencies in DNA damage repair limit the function of haematopoietic stem cells with age. *Nature* **447**, 725–729 (2007).
- Nijnik, A. *et al.* DNA repair is limiting for haematopoietic stem cells during ageing. *Nature* **447**, 686–690 (2007).
- Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. & Bonner, W. M. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J. Biol. Chem.* **273**, 5858–5868 (1998).
- Branzei, D. & Foiani, M. The DNA damage response during DNA replication. *Curr. Opin. Cell Biol.* **17**, 568–575 (2005).
- Furuta, T. *et al.* p21CDKN1A allows the repair of replication-mediated DNA double-strand breaks induced by topoisomerase I and is inactivated by the checkpoint kinase inhibitor 7-hydroxystaurosporine. *Oncogene* **25**, 2839–2849 (2006).
- Ishikawa, K., Ishii, H. & Saito, T. DNA damage-dependent cell cycle checkpoints and genomic stability. *DNA Cell Biol.* **25**, 406–411 (2006).
- Wang, Y. A., Elson, A. & Leder, P. Loss of p21 increases sensitivity to ionizing radiation and delays the onset of lymphoma in atm-deficient mice. *Proc. Natl Acad. Sci. USA* **94**, 14590–14595 (1997).
- Insinga, A. *et al.* Impairment of p53 acetylation, stability and function by an oncogenic transcription factor. *EMBO J.* **23**, 1144–1154 (2004).
- Bartkova, J. *et al.* Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. *Nature* **444**, 633–637 (2006).
- Di Micco, R. *et al.* Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. *Nature* **444**, 638–642 (2006).
- Paulson, Q. X. *et al.* Transgenic expression of E2F3a causes DNA damage leading to ATM-dependent apoptosis. *Oncogene* **27**, 4954–4961 (2008).
- Bartek, J., Bartkova, J. & Lukas, J. DNA damage signalling guards against activated oncogenes and tumour progression. *Oncogene* **26**, 7773–7779 (2007).
- Campisi, J. & d'Adda di Fagagna, F. Cellular senescence: when bad things happen to good cells. *Nature Rev. Mol. Cell Biol.* **8**, 729–740 (2007).
- Dimri, G. P. *et al.* A biomarker that identifies senescent human cells in culture and in aging skin *in vivo*. *Proc. Natl Acad. Sci. USA* **92**, 9363–9367 (1995).
- Speck, N. A. & Gilliland, D. G. Core-binding factors in haematopoiesis and leukaemia. *Nature Rev. Cancer* **2**, 502–513 (2002).
- Heyting, C., Huigen, A. & Den Engelse, L. Repair of ethylnitrosourea-induced DNA damage in the newborn rat. I. Alkali-labile lesions and *in situ* breaks. *Carcinogenesis* **1**, 769–778 (1980).
- Zhou, C. *et al.* DNA damage evaluated by gammaH2AX foci formation by a selective group of chemical/physical stressors. *Mutat. Res.* **604**, 8–18 (2006).
- Friedberg, E. C. Suffering in silence: the tolerance of DNA damage. *Nature Rev. Mol. Cell Biol.* **6**, 943–953 (2005).
- Linne, Z. DNA damage control by novel DNA polymerases: translesion replication and mutagenesis. *J. Biol. Chem.* **276**, 25639–25642 (2001).
- Guan, Y., Gerhard, B. & Hogge, D. E. Detection, isolation, and stimulation of quiescent primitive leukemic progenitor cells from patients with acute myeloid leukemia (AML). *Blood* **101**, 3142–3149 (2003).
- Ishikawa, F. *et al.* Chemotherapy-resistant human AML stem cells home to and engraft within the bone-marrow endosteal region. *Nature Biotechnol.* **25**, 1315–1321 (2007).
- Grignani, F. *et al.* High-efficiency gene transfer and selection of human hematopoietic progenitor cells with a hybrid EBV/retroviral vector expressing the green fluorescence protein. *Cancer Res.* **58**, 14–19 (1998).
- Minucci, S. *et al.* Oligomerization of RAR and AML1 transcription factors as a novel mechanism of oncogenic activation. *Mol. Cell* **5**, 811–820 (2000).
- Alcalay, M. *et al.* Acute myeloid leukemia fusion proteins deregulate genes involved in stem cell maintenance and DNA repair. *J. Clin. Invest.* **112**, 1751–1761 (2003).
- Insinga, A. *et al.* Inhibitors of histone deacetylases induce tumor-selective apoptosis through activation of the death receptor pathway. *Nature Med.* **11**, 71–76 (2005).
- van Os, R. *et al.* A limited role for p21Cip1/Waf1 in maintaining normal hematopoietic stem cell functioning. *Stem Cells* **25**, 836–843 (2007).
- Akashi, K., Traver, D., Miyamoto, T. & Weissman, I. L. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404**, 193–197 (2000).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank L. Luzi, D. Shing and M. Faretta for discussions and suggestions; M. Capillo, M. Stendardo and D. Sardella for assistance in maintenance of mouse colonies and for mouse genotyping; B. Amati, G. McVie, P. P. Di Fiore and C. Basilio for reviewing the manuscript; and P. Dalton for editing the manuscript. This study was supported by grants from the Italian Association on Cancer Research (AIRC), Ministero Italiano della Salute, Cariplo and the European Community (FP6: EPITRON and GENICA) to P.G.P. A.V. is a fellow of the Vollaro Foundation.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to A.V. ([andrea.viale@ifom-ieo-campus.it](mailto:andrea.viale@ifom-ieo-campus.it)) or P.G.P. ([piergiusseppe.pellicci@ifom-ieo-campus.it](mailto:piergiusseppe.pellicci@ifom-ieo-campus.it)).

## METHODS

**Mice.** PML-RAR knock-in mice<sup>10</sup> were provided by T. J. Ley and backcrossed in the C57BL/6J background (PR). p21<sup>-/-</sup> mice of mixed (C57BL/6 and 129Sv) background were obtained from Jackson laboratory and, for these studies, backcrossed in C57BL/6J or 129/Sv pure backgrounds (12 generations of backcrossing). PRp21<sup>-/-</sup> mice were generated by mating PR and p21<sup>-/-</sup> C57BL/6J mice. 129Sv p53<sup>-/-</sup> mice were obtained from the Jackson Laboratory. Mouse colonies were maintained in a certified animal facility in accordance with national guidelines.

**Limiting bone marrow transplantation.** Recipients C57BL/6 Ly5.1 mice (six to eight per group, per genotype) were lethally irradiated (10 Gy) the day before transplantation, injected with  $10^4$ ,  $10^5$  or  $5 \times 10^5$  Ly5.2<sup>+</sup> test bone marrow MNCs together with  $5 \times 10^5$  WT bone marrow MNCs (Ly5.1<sup>+</sup>), and analysed 4 months post-transplant for the presence of Ly5.2<sup>+</sup> in the PB. A recipient was defined as negative when less than 1% of PB leukocytes were Ly5.2<sup>+</sup>. CRU numbers were determined by Poisson statistical analysis (L-Calcul software, StemCell Technologies).

**Self-renewal determination.** Lethally irradiated recipients (Ly5.1<sup>+</sup>) were transplanted with stem cell equivalents of test and control bone marrow MNCs (after FACS evaluation of LT-HSC frequency). Eight weeks after transplantation, recipients were killed and the frequency of donor-derived (Ly5.2<sup>+</sup>) LT-HSC evaluated by FACS.

**Bone marrow competition.** One million test cells (Ly5.2<sup>+</sup>) (four to five per genotype) and  $10^6$  WT competitor cells (Ly5.1<sup>+</sup>) were pooled and injected into lethally irradiated C57BL/6 mice (five mice per genotype). Recipients' PB was analysed for donor contribution by flow-cytometric analysis. Re-populating unit (RU) = (percentage of test chimaerism)  $\times$  (number of competitor cells  $\times 10^{-5}$ ) / (100 - percentage of test chimaerism).

**Serial bone marrow transplantation.** Two million Ly5.2<sup>+</sup> bone marrow test MNCs were injected into lethally irradiated C57BL/6 Ly5.1<sup>+</sup> mice (six per genotype). Twelve to 16 weeks later, mice were killed, their bone marrow analysed for Ly5.2 donor contribution by flow cytometry and  $2 \times 10^6$  MNCs re-injected into irradiated recipients (Ly5.1<sup>+</sup>).

**LTC-IC assay.** Stromal layers were prepared from the WT bone marrow, cultured in 12.5% horse-serum IMDM supplemented with 12.5% FBS, penicillin, streptomycin, 1  $\mu$ M hydrocortisone and 50  $\mu$ M 2-mercaptoethanol at 33 °C. After 3 weeks, confluent stromal layers were trypsinized, irradiated (15 Gy) and subcultured in 6–12 multiwell plates. Cultures were then seeded with bone marrow MNCs ( $2.5 \times 10^5$  per well) or with infected and FACS-sorted Lin<sup>-</sup> ( $10^4$  per well) cells.

**CFC assay.** Cells derived from 5-week LTC-IC or  $5 \times 10^4$  bone marrow MNCs were seeded in methylcellulose (Stem Cell Technology) supplemented with 15% FCS, 2 ng ml<sup>-1</sup> IL3, 2 ng ml<sup>-1</sup> IL6, 50 ng ml<sup>-1</sup> SCF, 60 ng ml<sup>-1</sup> G-CSF and 20 ng ml<sup>-1</sup> GM-CSF (Peprotech). Colonies were scored after 10 days.

**DNA, RNA and protein analysis.** High molecular mass DNA was purified using the SNAP Whole Blood DNA Isolation Kit (Invitrogen). PCR analysis (GeneAmp, Applied Biosystems) of integration sites was performed on 200 ng DNA using the following primers: F-5'-AGGGACCCTATTGACGTTGAC-3' and R-5'-ACAGACAAAGCAAGGCTTGTAG-3' for PML-RAR; F-5'-GGCTG-GCAATGATGAAACT-3' and R-5'-CGCCATTCAAGGCTGTAGGAG-3' for AML1-ETO; F-5'-ACCACAGTCATGCCATCAC-3' and R-5'-TCCACCAC-CCTGTTGCTGTAG-3' for GAPDH. Total RNA was extracted using the Midi or Micro RNeasy kit (Qiagen) and retro transcribed with random hexamers (SuperScript First-Strand Synthesis System; Invitrogen). cDNA (6–8 ng) was used for RT-PCR, using the following primers: F-5'-CTGGGAGGGGA-

CAAGAG-3' and R-5'-GCTTGGAGTGATAGAAATCTG-3' for murine p21; F-5'-ACCTGCCAAGTATGATGAC-3' and R-5'-ACCACCCTGTGCTGT-AG-3' for murine GAPDH; F-5'-GACTCTCAGGGTCGAAAACG-3' and R-5'-GGATTAGGGCTTCCTCTTGG-3' for human p21; F-5'-ATCAGCAAT-GCCTCCTGCAC-3' and R-5'-TGGCATGGACTGTGGTCATG-3' for human GAPDH. Western blot antibodies:  $\alpha$ -p21 (C19, Santa Cruz),  $\alpha$ -RAR (C20, Santa Cruz),  $\alpha$ -HA (BAbCO) or  $\alpha$ -actin (AC-40, Sigma).

**Immunofluorescence.** Cells were deposited on charged slides (SuperFrost Plus, Menzel-Glaser), fixed (4% paraformaldehyde), permeabilized (0.1% Triton-X), blocked (3% BSA and 10% goat serum), stained with different antibodies ( $\alpha$ -p21: SX118, Dakocytomation;  $\alpha$ -phosphorylated Ser139 H2AX: Biolegend and Upstate) and counterstained with DAPI (Sigma) and Cy3- or Alexa488-conjugated secondary antibodies. Images were captured with a CCD (charge-coupled device) camera (Hamamatsu B/W CCD Camera CJ895), using a wide-field (Olympus BX61) or a confocal (Leica TCS SP2 AOBs) microscope with a  $\times 63/1.4$  NA oil immersion objective (HCX PL APO  $\times 63$  Lbd Bl Leica Microsystems).

## Flow cytometric analyses and sorting of haematopoietic sub-populations.

Donor contribution and multilineage reconstitution were assessed in PB samples by FACS analysis of the expression, respectively, of Ly5.1/Ly5.2 antigens or GR1, CD3 and B220 markers (eBioscience) in the Ly5.2<sup>+</sup> fraction. Homing of PML-RAR leukaemic blasts was investigated by injecting  $20 \times 10^6$  blasts (Ly5.2<sup>+</sup>) in the tail vein of Ly5.1<sup>+</sup> recipient mice. After 24 h, animals were killed and bone marrow mononuclear cells stained with  $\alpha$ -CD45Ly5.2. To analyse and isolate bone marrow LT-HSC, ST-HSC, MPP and CMP sub-populations<sup>11–13,18,46</sup>, MNCs, after blocking with 10% BSA and 10% rat serum, were stained with fluorochrome-conjugated antibodies against Sca1 (PE-CY5.5), c-Kit (APC), Flk-2 (PE), CD34 (FITC) or the lineage markers (Cd11b, GR1, Ter-119, IL7-R, CD3, CD4, CD8 and B220) (PE-CY7) (eBioscience). Samples were fixed with 2% formaldehyde (FA) for FACS analysis (FACSCanto, BD) or used fresh for cell sorting (FACSaria, BD). To stain leukaemic cells, we used a lineage-cocktail without myeloid markers, because Cd11b and GR1 are co-expressed in the c-Kit/Sca1 double positive blasts (unpublished observations). BrdU incorporation was analysed after four intraperitoneal injections (1 mg BrdU per mouse, every 6 h, into healthy and pre-leukaemic mice; mice killed 24 h after first injection<sup>17</sup>) or a single intraperitoneal injection (1 mg BrdU per mouse into overt leukaemic mice; mice killed 12 h after injection) using the FITC BrdU Flow Kit (BD). In this setting, surface staining was performed as described above except for the use of a biotinylated antibody against CD34 (eBioscience), detected by APC-CY7-conjugated streptavidin (eBioscience). To evaluate proliferation of LT-HSC, an equivalent number of sorted cells were cultured for 3–4 days in RPMI medium containing 10% FBS, 10 ng ml<sup>-1</sup> IL-3, IL-6, IL-11, SCF, TPO and FL3 ligand.

**5-FU treatment.** 5-FU (TEVA) was administered weekly intraperitoneally at a dose of 150 mg kg<sup>-1</sup>.

**PKH assay.** Bone marrow leukaemic blasts, stained with PKH-26 (Sigma) and FACS sorted, were injected in syngeneic recipients ( $10^7$  cells per mouse). When leukaemic mice were killed and the bone marrow PKH<sup>+</sup> cells FACS-sorted and re-injected into secondary recipients ( $10^3$  cells per mouse).

**Statistical analysis.** *In vitro* and *in vivo* data are presented as the mean  $\pm$  s.d. (standard error). Statistical analysis was performed using a two-tailed Student's *t*-test after the evaluation of variance. Results from survival experiments were analysed with a log-rank non-parametric test and expressed as Kaplan–Meier survival curves.



# WSTF regulates the H2A.X DNA damage response via a novel tyrosine kinase activity

Andrew Xiao<sup>1</sup>, Haitao Li<sup>2</sup>, David Shechter<sup>1</sup>, Sung Hee Ahn<sup>1</sup>, Laura A. Fabrizio<sup>3</sup>, Hediye Erdjument-Bromage<sup>3</sup>, Satoko Ishibe-Murakami<sup>2</sup>, Bin Wang<sup>4</sup>, Paul Tempst<sup>3</sup>, Kay Hofmann<sup>5</sup>, Dinshaw J. Patel<sup>2</sup>, Stephen J. Elledge<sup>4</sup> & C. David Allis<sup>1</sup>

**DNA double-stranded breaks present a serious challenge for eukaryotic cells. The inability to repair breaks leads to genomic instability, carcinogenesis and cell death. During the double-strand break response, mammalian chromatin undergoes reorganization demarcated by H2A.X Ser 139 phosphorylation ( $\gamma$ -H2A.X). However, the regulation of  $\gamma$ -H2A.X phosphorylation and its precise role in chromatin remodelling during the repair process remain unclear. Here we report a new regulatory mechanism mediated by WSTF (Williams–Beuren syndrome transcription factor, also known as BAZ1B)—a component of the WICH complex (WSTF–ISWI ATP-dependent chromatin-remodelling complex). We show that WSTF has intrinsic tyrosine kinase activity by means of a domain that shares no sequence homology to any known kinase fold. We show that WSTF phosphorylates Tyr 142 of H2A.X, and that WSTF activity has an important role in regulating several events that are critical for the DNA damage response. Our work demonstrates a new mechanism that regulates the DNA damage response and expands our knowledge of domains that contain intrinsic tyrosine kinase activity.**

One hallmark of the mammalian DNA double-strand break (DSB) response is the formation of ionizing-radiation-induced foci (IRIF), which are composed of compacted chromatin and numerous DNA repair and checkpoint proteins<sup>1,2</sup>. Despite considerable progress in understanding the signalling pathways leading to checkpoint control and DNA repair, the nature of the specialized chromatin structures at IRIF is not well understood. One of the earliest events occurring at IRIF is the phosphorylation of H2A.X (also known as H2afx), a specialized histone H2A variant, at Ser 139 (referred to as  $\gamma$ -H2A.X) by the ATM and ATR kinases<sup>3</sup>. H2A.X-deficient mouse embryonic fibroblasts (MEFs) and B and T cells show pronounced levels of genomic instability<sup>4</sup>. Class-switch recombination and spermatogenesis are also defective in H2A.X-deficient mice, further indicating its involvement in DNA damage repair<sup>4–6</sup>. Moreover, H2A.X deficiency accelerates B and T cell lymphoma development in p53-deficient mice<sup>5,7</sup>. Consistent with these functions in mammalian cells, phosphorylation of the equivalent site on yeast H2A (Ser 129) is found at DSB sites and spreads to ~50 kilobase (kb) of the flanking regions<sup>8</sup>. In mammals, this phosphorylation event directly recruits Mdc1 and sets in motion the recruitment of other factors such as 53BP1 (also known as Trp53bp1), Rnf8 and the Brca1 A complex<sup>9</sup>. Furthermore, several recent studies have also indicated that ATP-dependent chromatin remodelling complexes are engaged in DNA repair pathways at these sites. The yeast NuA4 and INO80 complexes, for example, are recruited to the damaged chromatin by  $\gamma$ -H2A.X<sup>10–12</sup>.

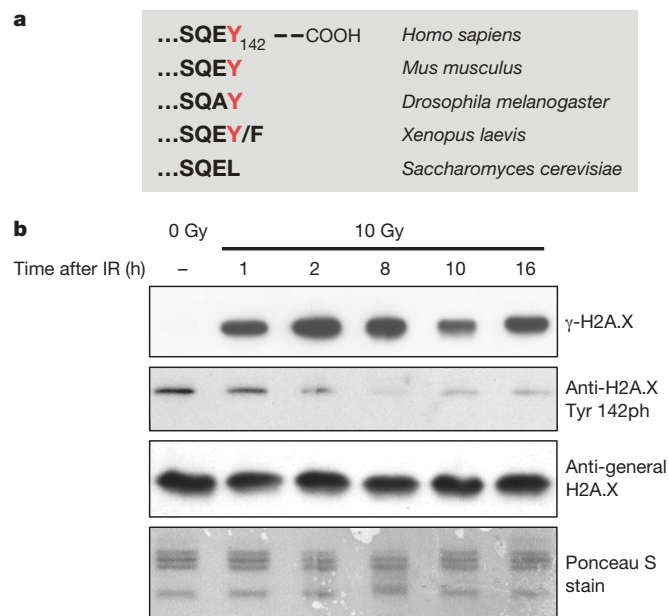
Mammalian H2A.X bears several differences with lower eukaryotes. For example, H2A.X is a minor H2A variant in mammalian cells (1–10%, ref. 3), whereas the main yeast form of H2A is most similar to H2A.X because it contains the signature carboxy-terminal

sequence of mammalian H2A.X<sup>3</sup>. In mammalian cells, the loss of SWI/SNF chromatin remodelling complex expression leads to defects in the H2A.X DNA damage response. However, it is unclear whether these defects are due to the lack of direct regulation by the SWI/SNF complex or to other indirect pathways<sup>13</sup>. Given the unique features of mammalian H2A.X, we sought to find new factors that are directly involved in regulating H2A.X. Our results define a novel DNA damage response pathway regulating H2A.X function, mediated through the WSTF–SNF2H chromatin remodelling complex and the phosphorylation of H2A.X at Tyr 142 in mammals. Unexpectedly, we determined that the amino-terminal domain of WSTF, including its WAC domain, exhibits tyrosine kinase activity towards Tyr 142 of H2A.X. We further show that the tyrosine kinase activity of WSTF is required for eliciting a number of critical molecular events during the DNA damage response in mammalian cells.

## Regulation of Tyr 142 phosphorylation

A role in the DNA damage response for mammalian H2A.X ( $\gamma$ -H2A.X) is well documented, although its regulation and underlying mechanism of action is only partially understood (reviewed in ref. 3). Further inspection of the C terminus of H2A.X showed a tyrosine (Tyr 142 in mammals) that exists in metazoans, but is absent in unicellular eukaryotes such as yeast (Fig. 1a). Notably, two forms of H2A.X exist in the *Xenopus* genome, which are different at this residue (the 'F' and the 'Y' form, Fig. 1a), are differentially expressed during development (D.S. *et al.*, manuscript in preparation). Although recent studies suggest a role for Tyr 142 in recruiting Mdc1 (refs 14–19) the *in vivo* function of Tyr 142, especially in the regulation of  $\gamma$ -H2A.X, remains unclear<sup>17</sup>. We proposed that Tyr 142 might be phosphorylated in H2A.X under certain physiological conditions (Fig. 1b).

<sup>1</sup>Laboratory of Chromatin Biology, The Rockefeller University, New York, New York 10065, USA. <sup>2</sup>Structural Biology Program, <sup>3</sup>Molecular Biology Program, Memorial-Sloan-Kettering Cancer Center, New York, New York 10065, USA. <sup>4</sup>Howard Hughes Medical Institute, Department of Genetics, Harvard Partners Center for Genetics and Genomics, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>5</sup>Miltenyi Biotec GmbH, 50829 Koeln, Germany.



**Figure 1 | Tyr142 of H2A.X is a new phosphorylation mark regulated by DNA damage signals.** **a**, Comparison of the extreme C-terminal sequence of H2A.X demonstrates that Tyr 142 is conserved in metazoans (mammals, frogs and fruitflies) but not in unicellular eukaryotes. **b**, Primary MEF cells were treated with 10 Gy of ionizing radiation (IR) and recovered for a period of time as indicated. Acid-extracted histones were separated by SDS-PAGE and subjected to immunoblotting. H2A.X Tyr 142 phosphorylation levels in MEFs gradually decline, reaching a minimum at 8 h. The  $\gamma$ -H2A.X signal is initiated on damage and maintained up to 16 h after ionizing radiation.

Preliminary studies using a pan anti-phosphotyrosine antibody indicated that H2A.X was phosphorylated before DNA damage (data not shown). To investigate further whether Tyr 142 is indeed phosphorylated, an antibody raised against an extreme C-terminal peptide containing phosphorylated Tyr 142 of H2A.X was generated and shown to be highly selective for H2A.X Tyr 142 phosphorylation (hereafter termed anti-H2A.X Tyr 142(ph); Supplementary Fig. 1). In MEF cells, Tyr 142 is constitutively phosphorylated under normal growth conditions and becomes gradually dephosphorylated during the DNA damage response, while  $\gamma$ -H2A.X increases (Fig. 1b). As it has been reported that H2A.X is highly enriched in *Xenopus* oocytes and eggs<sup>20,21</sup>, we addressed whether its Tyr 142 phosphorylation status would be altered in response to DNA damage treatment, in the context of *Xenopus* early embryonic development. Similar to mammalian cells, Tyr 142 phosphorylation levels are greatly decreased in response to DNA damage (Supplementary Fig. 1a).

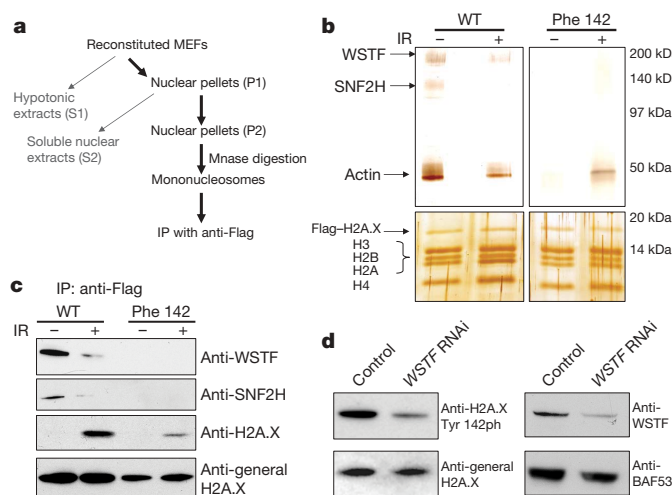
Because the major form of H2A in yeast carries the 'SQEL' motif (Fig. 1a and ref. 3), we examined whether a point mutant mimicking the mammalian 'SQEY' motif would constitute a new phospho-acceptor. Similar to the mammalian cells, this mutant yeast strain (Leu132Tyr) is strongly reactive to anti-H2A.X Tyr 142(ph) antibodies with reactivity greatly diminished after DNA damage treatment (Supplementary Fig. 1b). Collectively, these results raise the intriguing possibility that an evolutionarily conserved DNA-damage-regulated phosphotyrosine pathway exists and regulates H2A.X and probably other DNA damage responsive proteins.

### The WSTF-SNF2H complex interacts with H2A.X

To investigate the mechanisms that regulate H2A.X function, we sought to isolate protein complexes directly associated with H2A.X *in vivo*. Primary H2A.X<sup>-/-</sup> MEF cells were reconstituted with H2A.X constructs (wild type or Tyr142Phe mutant), with or without an N-terminal Flag epitope tag. The Flag tag did not interfere with typical DNA damage response pathways, including  $\gamma$ -H2A.X (data not

shown). Three independent cell lines (wild type and mutants) were constructed in which the expression levels were comparable (50–80%) to the control primary MEFs (with two intact endogenous H2A.X alleles) and most of the reconstituted cells (~90%) have similar expression levels (data not shown). To enrich for proteins or complexes that may associate with H2A.X mononucleosomes, we adapted an approach that successfully enriched such chromatin particles for subsequent biochemical purifications<sup>22</sup>. This method, followed by immunoprecipitation with anti-Flag antibodies, greatly enriched H2A.X-containing mononucleosomes and associated proteins (Fig. 2a and Supplementary Fig. 2). As shown by silver staining, a small number of polypeptides besides histones were associated with the wild-type H2A.X mononucleosomes at a stoichiometric level, the amount of which was decreased after cells were treated with ionizing radiation (Fig. 2b). In addition, these interactions are reduced with Tyr142Phe mutant H2A.X mononucleosomes (Fig. 2b). Subsequent mass spectrometry analysis demonstrated that two of the more prominent proteins are SNF2H (also known as SMARCA5; 140 kDa), a mammalian homologue of the ISWI ATPase<sup>23,24</sup>, and WSTF (171 kDa), a member of the BAZ/WAL family chromatin remodelling factors (Supplementary Fig. 3)<sup>25–27</sup>. These proteins constitute the WICH complex (WSTF-ISWI ATP-dependent chromatin-remodelling complex), which mobilizes nucleosomes *in vitro* and is suggested to be involved in the regulation of DNA replication<sup>28,29</sup>.

We also investigated if other protein factors present in ISWI-containing chromatin remodelling complexes, such as CHRAC15 (also known as CHRAC1) and CHRAC17 (POLE3) (ref. 30), co-purified with H2A.X-containing mononucleosomes at sub-stoichiometric levels. However, none of these components were detected by either mass spectrometry or immunoblotting approaches (data not shown), leading us to the tentative conclusion that only the WICH complex copurifies with H2A.X-containing mononucleosomes at



**Figure 2 | The WSTF-SNF2H chromatin remodelling complex is specifically associated with H2A.X nucleosomes *in vivo*.** **a**, Purification scheme of H2A.X-containing mononucleosomes. MEFs reconstituted with Flag-H2A.X (wild type or Tyr142Phe) were treated with or without ionizing radiation (IR). **b**, Two polypeptides migrating at 145 and 171 kDa were associated with undamaged wild type (WT, left) but not the Tyr142Phe mutant (right) H2A.X-mononucleosomes in a silver-stained gel. Mass spectrometry analyses showed that these polypeptides were WSTF (171 kDa) and SNF2H (145 kDa). The third band at 50 kDa was identified by mass spectrometry as  $\beta$ -actin. Staining and mass spectrometry of the lower molecular weight bands (<20 kDa) showed similar levels of H2A.X and other core histones. **c**, The association between the WICH complex and undamaged wild-type H2A.X mononucleosomes was confirmed by immunoprecipitation (IP) and western blotting experiments. **d**, In WSTF RNAi cells, the expression level of WSTF was greatly diminished, and the H2A.X Tyr 142 phosphorylation level was greatly reduced.



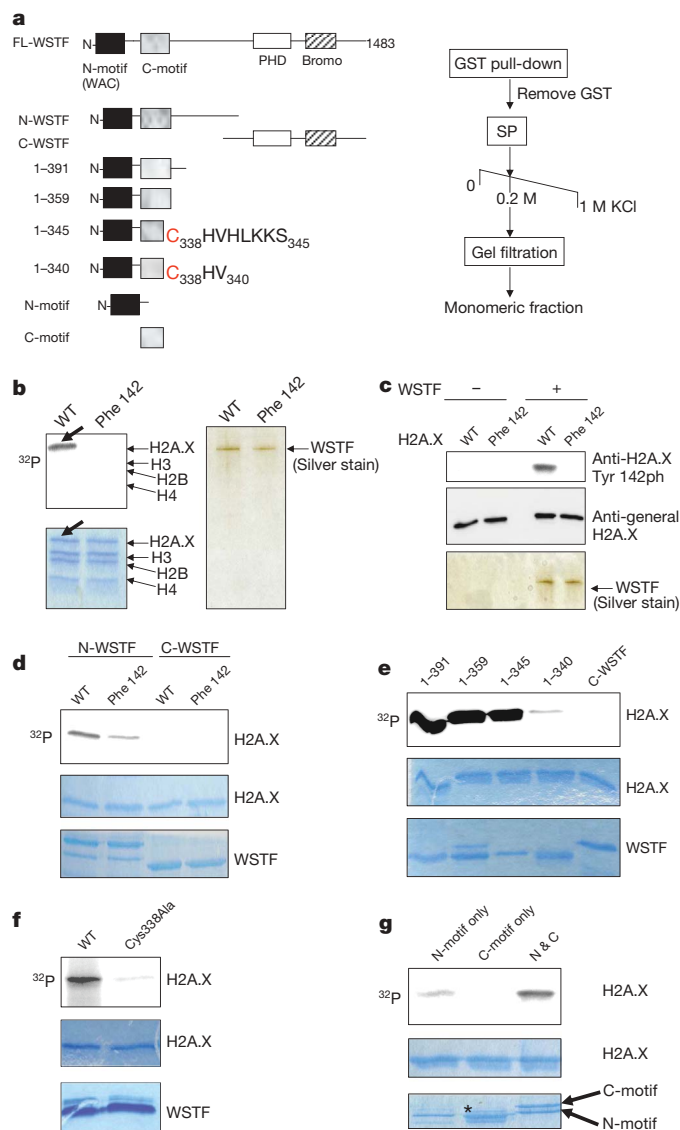
stoichiometric levels *in vivo*. We next confirmed these association results by immunoblot analyses of the complexes co-immunoprecipitated with H2A.X-containing nucleosomes. Consistent with the mass spectrometry results, WSTF and SNF2H were enriched in these complexes (Fig. 2c).

To investigate the roles of WSTF in regulating the function of H2A.X, several independent WSTF knockdown NIH3T3 cell lines (WSTF RNA interference (RNAi) cells) were generated by infection with a short hairpin WSTF RNAi construct, in which the WSTF protein expression is greatly reduced (Fig. 2d). Additionally, we attempted to knockdown SNF2H with a similar approach, but SNF2H knockdown resulted in S-phase defects and cell death (data not shown) as described in SNF2H-knockout mice studies<sup>31</sup>; presumably, SNF2H functions in other chromatin remodelling complexes that are involved in a variety of vital cell functions<sup>24</sup>. The specificity of the RNAi knockdown approach was demonstrated by rescuing WSTF expression with the human WSTF gene (the sequences in the WSTF messenger RNA targeted by the RNAi construct are different between human and mouse; see later). Notably, WSTF deficiency leads to a decrease in Tyr 142 phosphorylation (Fig. 2d), indicating that WSTF may be also directly involved in regulating Tyr 142 phosphorylation *in vivo* (see below). These data suggest a regulatory mechanism of H2A.X that involves the WICH complex and Tyr 142 phosphorylation.

### WSTF is a tyrosine kinase that phosphorylates Tyr 142 of H2A.X

To investigate further the function of WSTF, we purified recombinant, full-length human WSTF proteins in insect cells (Fig. 3a). A single protein band migrating at the expected molecular weight (171 kDa) was detected by silver staining (Fig. 3b). Mass spectrometry analysis on the purified samples confirmed that this protein was WSTF, and did not detect any other proteins except trace amounts of heat shock proteins (data not shown). The recombinant full-length WSTF protein phosphorylated H2A.X-containing nucleosomes (Fig. 3b) that were reconstituted *in vitro*, using ATP and divalent manganese ion  $Mn^{2+}$  (but not  $Mg^{2+}$ , see Supplementary Fig. 4) as cofactors. To address whether WSTF can specifically phosphorylate Tyr 142 of H2A.X, we also reconstituted Tyr142Phe mutant nucleosomes. The Tyr142Phe mutant does not affect histone octamer or nucleosome formation during reconstitution (data not shown). In *in vitro* kinase assays, WSTF phosphorylated the wild-type but not Tyr142Phe H2A.X, indicating that Tyr 142 is the main site of phosphorylation (Fig. 3b). The *in vitro* kinase activity of WSTF towards Tyr 142 of H2A.X was confirmed by immunoblotting using anti-H2A.X Tyr 142(ph) antibodies (Fig. 3c).

Because no detectable proteins other than trace amounts of heat shock proteins co-purified with WSTF from insect cells (Fig. 3b), we proposed that WSTF has an intrinsic kinase activity. However, WSTF does not share homology with known kinase domains. Therefore we generated truncated recombinant WSTF proteins (Fig. 3a) to investigate which portion of the protein is required for this kinase activity. Using this approach, we demonstrated that the N-terminal portion of WSTF is necessary and sufficient for kinase activity, strongly indicating that the kinase motif resides in its N-terminal portion (Fig. 3d). Additional structure function analyses determined that residues 1–345 contained the active kinase domain composed of a WAC domain and a newly identified C-motif that are both required for optimal kinase activity (Fig. 3e, g and Methods). A Cys338Ala mutation generated a kinase dead allele that maintained proper folding (Fig. 3f, Supplementary Fig. 6 and Methods). To confirm the intrinsic activity of the putative WSTF kinase domain further, we generated recombinant WSTF proteins in *Escherichia coli*, because protein kinases, especially tyrosine kinases, are rare in the *E. coli* genome. Bioinformatic and biochemical analysis of the kinase domain of WSTF reveals two well-structured motifs (N-motif, containing the conserved WAC domain<sup>27</sup>, and a previously unreported C-motif; Fig. 3a and Supplementary Fig. 7). Co-expression of both motifs



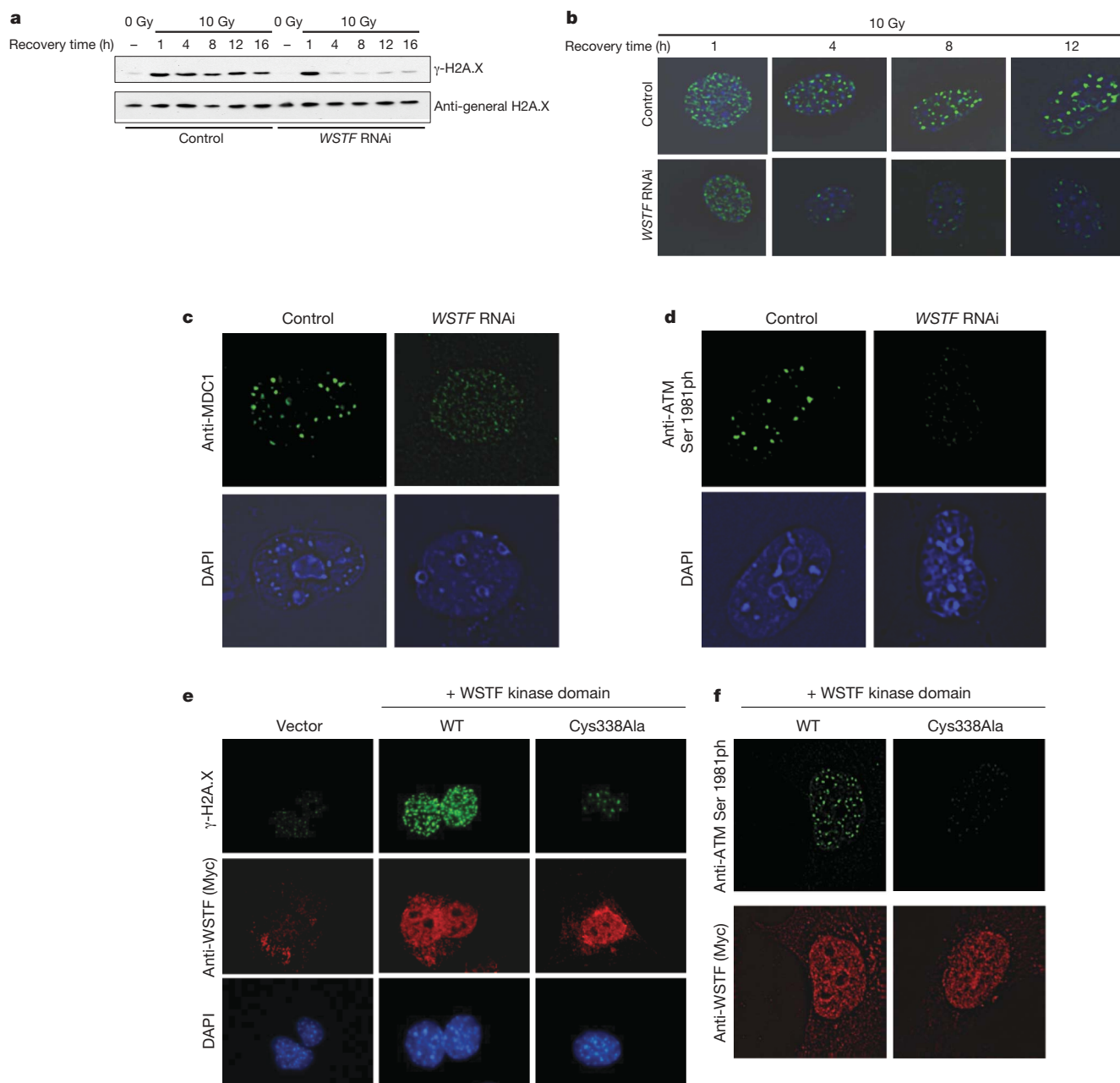
**Figure 3 | WSTF contains a kinase domain that phosphorylates Tyr 142 of H2A.X.** **a**, Schematic demonstration of the domain architecture of the human WSTF protein and a series of recombinant proteins representing portions of WSTF. Bromo, bromo domain; FL, full-length; GST, glutathione S-transferase; PHD, PHD (plant homeodomain) finger domain; SP, SP sepharose chromatography. **b–f**, Recombinant WSTF proteins were generated in insect cells; the N- and C-motifs (**g**) were generated in *E. coli*. **b**, Tyr 142 in the nucleosome is phosphorylated by recombinant full-length WSTF protein. WT, wild type. **c**, The specific phosphorylation of H2A.X at Tyr 142 was detected by immunoblotting with anti-H2A.X Tyr 142(ph) antibodies. In **d–g**, free H2A.X proteins were used as substrates in *in vitro* kinase assays. **d**, N-WSTF, but not C-WSTF, had kinase activity. **e**, The WSTF 1–340 construct had much reduced kinase activity (<50-fold) in comparison to the other constructs. **f**, The Cys338Ala point mutant (derived from wild-type 1–359) had much reduced kinase activity. This mutation does not change the global folding properties of the kinase domain (Supplementary Fig. 6). **g**, The co-expressed N-motif and C-motif had enhanced kinase activity towards H2A.X, whereas the N-motif alone had minimal kinase activity. The C-motif construct, when expressed alone, was unstable and partially degraded from its C terminus, however, it was protected when co-expressed with the N-motif.

restores the kinase activity, whereas the N-terminal motif alone had a much reduced activity (Fig. 3g). These results strongly indicate that the N- and C-motif are required for the optimal kinase activity. Collectively, our results demonstrate that WSTF has a previously unidentified intrinsic kinase activity by means of a novel kinase domain.

### WSTF has a critical role in foci formation

Because the functions of the WICH complex, especially that of WSTF, in the DNA damage response have not been explored, we investigated the role of WSTF in the DNA damage response. On DNA damage treatment, the initial  $\gamma$ -H2A.X phosphorylation of H2A.X and its foci formation in *WSTF* RNAi cells (as described in Fig. 2) were similar to control cells (up to 1 h after ionizing radiation, Fig. 4a). However, in control cells, the  $\gamma$ -H2A.X level remained unchanged until 16 h after ionizing radiation. In contrast, the level of  $\gamma$ -H2A.X phosphorylation rapidly declined in *WSTF* RNAi cells 4 h after ionizing radiation treatment (Fig. 4a). It is well-established that the number and

morphology of  $\gamma$ -H2A.X foci undergo notable changes during the DNA damage response. Initially, a larger number of small foci are formed, whereas at late stages of the DNA damage response, only a few large foci are usually observed<sup>32</sup>. As expected, large  $\gamma$ -H2A.X foci were formed in control cells starting 4 h after ionizing radiation (Fig. 4b, top panels) and persisting until 12 h, whereas the overall level of  $\gamma$ -H2A.X phosphorylation remains relatively constant. However, this morphological progression was not observed in *WSTF* RNAi cells despite the initial formation of small  $\gamma$ -H2A.X speckles as in the controls (Fig. 4b, bottom panels). Instead, the amount and intensity of the  $\gamma$ -H2A.X foci were much less than in controls (only 16% of the *WSTF*



**Figure 4 | WSTF is critical for the maintenance of  $\gamma$ -H2A.X phosphorylation after DNA damage.** **a**,  $\gamma$ -H2A.X maintenance is defective in *WSTF* RNAi cells after DNA damage treatment. The time points labelled indicate the recovery time after the ionizing radiation treatment. **b**, Immunofluorescent staining experiments were performed on control and *WSTF* RNAi cells fixed at different time points (as labelled) following 10 Gy of ionizing radiation treatment. **c**, **d**, Mdc1 and phos-ATM recruitment are also defective in *WSTF*-deficient cells. Immunofluorescent staining experiments were performed with anti-ATM Ser 1981(ph) and anti-Mdc1

antibodies on control or *WSTF* RNAi cells, 8 h after 10 Gy of ionizing radiation treatment. DAPI, 4,6-diamidino-2-phenylindole. **e**, *WSTF* RNAi cells were complemented with the wild-type (WT) or mutant (Cys338Ala) kinase domain constructs (1–359, Myc-epitope-tagged at the N terminus) of *WSTF*. Eight hours after treated with 10 Gy of ionizing radiation, cells were fixed and co-stained with  $\gamma$ -H2A.X (green) and anti-Myc antibodies (red). **f**, The kinase activity of *WSTF* (red) is also critical for phos-ATM foci maintenance (green). Original magnification **b–f**,  $\times 60$ .



RNAi retain  $\gamma$ -H2A.X foci 4 h after ionizing radiation treatment), and no large foci were observed in *WSTF* RNAi cells during the entire experiment, even 12 h after ionizing radiation (Fig. 4b). Because the major kinases for  $\gamma$ -H2A.X phosphorylation are ATM and ATR, and  $\gamma$ -H2A.X foci maintenance is dependent on sustained recruitment of active ATM to the damage foci via Mdc1 (ref. 17), we addressed whether the relocation of ATM and Mdc1 were defective in *WSTF*-deficient cells. In control cells, large Mdc1 foci were observed at a late stage of the DNA damage response (8 h after 10 Gy of ionizing radiation; Fig. 4c), whereas similar foci were not observed in the *WSTF*-deficient cells, indicating that the recruitment of Mdc1 is defective at a late stage of the DNA damage response in these cells. Autophosphorylation at Ser 1981 (referred to as phos-ATM hereafter) is one of the indicators of ATM activation during the DNA damage response<sup>33–35</sup>. Foci formation of phos-ATM (and potentially other targets of ATM) was also impaired in *WSTF*-deficient cells (Fig. 4d). These data strongly suggest that *WSTF* has a critical role in the recruitment of active ATM and Mdc1 to the damage sites, both of which are critical for  $\gamma$ -H2A.X foci formation.

We next investigated whether the kinase activity of *WSTF* is required for its function during the DNA damage response, namely the maintenance of phos-ATM and  $\gamma$ -H2A.X foci. Because the *WSTF* mRNA sequences targeted by the short hairpin RNAi constructs are different between human and mouse (see earlier), we complemented *WSTF* RNAi cells (derived from mouse 3T3 cells) with wild-type or Cys338Ala mutant kinase domain constructs of human *WSTF*. This approach successfully restored the expression of kinase domain of *WSTF* (Fig. 4e, f). The expression of the wild-type *WSTF* kinase domain rescued the  $\gamma$ -H2A.X and phos-ATM foci formation defects in the *WSTF* RNAi cells (Fig. 4e, f). In nearly 100% of the cells complemented with the wild-type construct,  $\gamma$ -H2A.X and phos-ATM foci were observed at least 8 h after DNA damage treatment (Fig. 4e, f). Conversely, the phos-ATM and  $\gamma$ -H2A.X foci formation were as defective in cells expressing the Cys338Ala mutant as in *WSTF* RNAi cells containing vector alone (Fig. 4e, f). Taken together, these results demonstrate that *WSTF* has an important involvement in the DNA damage response by means of its kinase activity—an activity that targets H2A.X Tyr 142 and possibly other yet unknown substrates.

We wondered, however, whether Ser 139 phosphorylation was required for the steady-state balance of H2A.X phosphorylation at Tyr 142, or vice versa. As previously reported<sup>15,18</sup>, the maintenance of  $\gamma$ -H2A.X phosphorylation is affected in the *Mdc1*<sup>−/−</sup> MEF cells (Supplementary Fig. 8). Tyr 142, however, becomes gradually dephosphorylated in the *Mdc1*<sup>−/−</sup> cells as in the wild-type control cells (Supplementary Fig. 8), indicating that sustained  $\gamma$ -H2A.X phosphorylation is not required for Tyr 142 dephosphorylation. Conversely, the Ser 139 phosphorylation level and foci formation were greatly reduced after DNA damage in both Tyr142Leu and Tyr142Phe mutant cells (Supplementary Fig. 8). In addition, mutations on Tyr 142 affect Mdc1 binding to the H2A.X C-tail phosphorylated at Ser 139 (Supplementary Fig. 8). These data suggest that these phosphorylation events may be coordinated during the DNA damage response (see later in the text).

## Discussion

Our studies highlight a new regulatory mechanism controlling histone H2A.X function mediated by the *WSTF*–SNF2H (WICH) chromatin remodelling complex. We found a previously unknown phosphorylation site on H2A.X, Tyr 142, and demonstrated that it has a vital role in the DNA damage response. We also discovered that the N-terminal domain of *WSTF* has a previously unrecognized tyrosine kinase activity for this site (see below). *WSTF*, a gene frequently deleted in human Williams–Beuren (also known as Williams) syndrome, has previously been shown to be the regulatory component of the WICH chromatin remodelling complex<sup>25,29</sup>. Our findings have

therefore identified an unexpected link between a histone minor variant, H2A.X, a histone covalent modification and ATP-dependent chromatin remodelling mechanisms that control the mammalian DNA damage response.

The kinase domain of *WSTF* is composed of two putative motifs. The N-motif contains the highly conserved WAC domain, which is found in many eukaryotic proteins, whereas the C-motif is much more divergent. Besides H2A.X, *WSTF* may have other downstream targets given its multiple roles in DNA damage repair, replication and transcription. Williams–Beuren syndrome is a well-documented genetic disorder characterized by developmental defects and clinical behavioural phenotypes<sup>36</sup>, so identifying the link between the biochemical function of *WSTF* and the clinical manifestation of Williams–Beuren syndrome remains a challenge for future studies. In addition, *WSTF* kinase activity and Tyr 142 phosphorylation may have several roles during DNA damage repair (see Supplementary Information for a detailed discussion). First, pre-existing phosphorylated Tyr 142 may be needed at the break to adjust local chromatin structure for the later maintenance of Ser 139 phosphorylation. Subsequent dephosphorylation of Tyr 142 after DNA damage could act to enhance Mdc1 and ATM recruitment to extend and maintain  $\gamma$ -H2A.X phosphorylation. Second, *WSTF* may regulate  $\gamma$ -H2A.X foci maintenance through other pathways and unidentified targets, which may be involved in recruiting or stabilizing Mdc1 and ATM at damage foci. Finally, in the absence of Tyr 142 phosphorylation, the kinetics of the phosphorylation/dephosphorylation cycle of  $\gamma$ -H2A.X may be altered such that both the phosphorylation and dephosphorylation of H2A.X on Ser 139 occur much more rapidly, and the foci form and disassemble in less than 4 h. Identification of the signalling pathways using H2A.X Tyr 142 phosphorylation or other physiologically relevant phosphorylation events regulated by the WICH chromatin remodelling complex will be of interest in future studies.

## METHODS SUMMARY

**Plasmids and cell culture.** All plasmid constructs were generated with standard protocols (see Methods). Recombinant MSCV virus carrying H2A.X or *WSTF* constructs were packaged in Phoenix cells and used to infect MEFs or NIH3T3 cells. Murine stem cell virus (MSCV) production and infection followed standard protocols<sup>37</sup>. For each construct, three independent lines were derived. The expression levels were checked by immunoblots and immunofluorescence.

**Purification of H2A.X-containing mononucleosomes and associated protein factors.** MEF fractionation and chromatin pellet isolation were performed as described<sup>22</sup>. Chromatin pellets were briefly digested with MnaI (Sigma), and the generation of mononucleosomes was monitored by electrophoresis<sup>38</sup> and subject to mass spectrometry analysis.

**Generation of recombinant *WSTF* proteins.** Recombinant baculovirus carrying the full-length *WSTF* (Flag-epitope-tagged at C terminus) was produced using the BaculoGold-BEVS kit (BD Biosciences). The full-length *WSTF* proteins were purified using a published protocol<sup>25</sup>. The viruses that express truncated *WSTF* constructs were produced with the Bac-to-Bac (Invitrogen) approach and purified (Methods). The N- and C-motif of the *WSTF* kinase domain were generated in *E. coli* (Methods).

**In vitro kinase assay conditions.** Recombinant full-length (50–100 ng) or truncated *WSTF* proteins (0.1–0.5  $\mu$ g) were incubated with free histone proteins (1–2  $\mu$ g), histone octamers or nucleosomes assembled *in vitro* (0.5–1  $\mu$ g) in 20 mM Tris, pH 7.4, and 150 mM NaCl supplemented with <sup>32</sup>P- $\gamma$ -ATP (NEN) and 1 mM MnCl<sub>2</sub>, at 25 °C for 45 min.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 28 July; accepted 27 November 2008.

Published online 17 December 2008.

1. Khanna, K. K. & Jackson, S. P. DNA double-strand breaks: signaling, repair and the cancer connection. *Nature Genet.* **27**, 247–254 (2001).
2. Zhou, B. B. & Elledge, S. J. The DNA damage response: putting checkpoints in perspective. *Nature* **408**, 433–439 (2000).
3. Redon, C. *et al.* Histone H2A variants H2AX and H2AZ. *Curr. Opin. Genet. Dev.* **12**, 162–169 (2002).

4. Celeste, A. *et al.* Genomic instability in mice lacking histone H2AX. *Science* **296**, 922–927 (2002).
5. Celeste, A. *et al.* H2AX haploinsufficiency modifies genomic stability and tumor susceptibility. *Cell* **114**, 371–383 (2003).
6. Reina-San-Martin, B. *et al.* H2AX is required for recombination between immunoglobulin switch regions but not for intra-switch region recombination or somatic hypermutation. *J. Exp. Med.* **197**, 1767–1778 (2003).
7. Bassing, C. H. *et al.* Histone H2AX: a dosage-dependent suppressor of oncogenic translocations and tumors. *Cell* **114**, 359–370 (2003).
8. Unal, E. *et al.* DNA damage response pathway uses histone modification to assemble a double-strand break-specific cohesin domain. *Mol. Cell* **16**, 991–1002 (2004).
9. Harper, J. W. & Elledge, S. J. The DNA damage response: ten years after. *Mol. Cell* **28**, 739–745 (2007).
10. Morrison, A. J. *et al.* INO80 and  $\gamma$ -H2AX interaction links ATP-dependent chromatin remodeling to DNA damage repair. *Cell* **119**, 767–775 (2004).
11. Downs, J. A. *et al.* Binding of chromatin-modifying activities to phosphorylated histone H2A at DNA damage sites. *Mol. Cell* **16**, 979–990 (2004).
12. van Attikum, H., Fritsch, O., Hohn, B. & Gasser, S. M. Recruitment of the INO80 complex by H2A phosphorylation links ATP-dependent chromatin remodeling with DNA double-strand break repair. *Cell* **119**, 777–788 (2004).
13. Park, J. H. *et al.* Mammalian SWI/SNF complexes facilitate DNA double-strand break repair by promoting  $\gamma$ -H2AX induction. *EMBO J.* **25**, 3986–3997 (2006).
14. Stewart, G. S., Wang, B., Bignell, C. R., Taylor, A. M. & Elledge, S. J. MDC1 is a mediator of the mammalian DNA damage checkpoint. *Nature* **421**, 961–966 (2003).
15. Goldberg, M. *et al.* MDC1 is required for the intra-S-phase DNA damage checkpoint. *Nature* **421**, 952–956 (2003).
16. Lou, Z., Minter-Dykhouse, K., Wu, X. & Chen, J. MDC1 is coupled to activated CHK2 in mammalian DNA damage response pathways. *Nature* **421**, 957–961 (2003).
17. Stucki, M. *et al.* MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell* **123**, 1213–1226 (2005).
18. Lou, Z. *et al.* MDC1 maintains genomic stability by participating in the amplification of ATM-dependent DNA damage signals. *Mol. Cell* **21**, 187–200 (2006).
19. Lee, M. S., Edwards, R. A., Thede, G. L. & Glover, J. N. Structure of the BRCT repeat domain of MDC1 and its specificity for the free COOH-terminal end of the  $\gamma$ -H2AX histone tail. *J. Biol. Chem.* **280**, 32053–32056 (2005).
20. Dimitrov, S., Dasso, M. C. & Wolffe, A. P. Remodeling sperm chromatin in *Xenopus laevis* egg extracts: the role of core histone phosphorylation and linker histone B4 in chromatin assembly. *J. Cell Biol.* **126**, 591–601 (1994).
21. Kleinschmidt, J. A. & Steinbeisser, H. DNA-dependent phosphorylation of histone H2A.X during nucleosome assembly in *Xenopus laevis* oocytes: involvement of protein phosphorylation in nucleosome spacing. *EMBO J.* **10**, 3043–3050 (1991).
22. Mendez, J. & Stillman, B. Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis. *Mol. Cell Biol.* **20**, 8602–8612 (2000).
23. Fyodorov, D. V. & Kadonaga, J. T. The many faces of chromatin remodeling: SWItching beyond transcription. *Cell* **106**, 523–525 (2001).
24. Becker, P. B. & Horz, W. ATP-dependent nucleosome remodeling. *Annu. Rev. Biochem.* **71**, 247–273 (2002).
25. Ito, T. *et al.* ACF consists of two subunits, Acl1 and ISWI, that function cooperatively in the ATP-dependent catalysis of chromatin assembly. *Genes Dev.* **13**, 1529–1539 (1999).
26. Poot, R. A. *et al.* HuCHRAC, a human ISWI chromatin remodelling complex contains hACF1 and two novel histone-fold proteins. *EMBO J.* **19**, 3377–3387 (2000).
27. Jones, M. H., Hamana, N., Nezu, J. & Shimane, M. A novel family of bromodomain genes. *Genomics* **63**, 40–45 (2000).
28. Poot, R. A. *et al.* The Williams syndrome transcription factor interacts with PCNA to target chromatin remodelling by ISWI to replication foci. *Nature Cell Biol.* **6**, 1236–1244 (2004).
29. Bozhenok, L., Wade, P. A. & Varga-Weisz, P. WSTF-ISWI chromatin remodeling complex targets heterochromatic replication foci. *EMBO J.* **21**, 2231–2241 (2002).
30. Kukimoto, I., Elderkin, S., Grimaldi, M., Oelgeschlager, T. & Varga-Weisz, P. D. The histone-fold protein complex CHRAC-15/17 enhances nucleosome sliding and assembly mediated by ACF. *Mol. Cell* **13**, 265–277 (2004).
31. Stopka, T. & Skoultchi, A. I. The ISWI ATPase Snf2h is required for early mouse development. *Proc. Natl Acad. Sci. USA* **100**, 14097–14102 (2003).
32. Deltre, G. & Bazett-Jones, D. P. Beyond repair foci: subnuclear domains and the cellular response to DNA damage. *Cell cycle* **6**, 1864–1872 (2007).
33. Bakkenist, C. J. & Kastan, M. B. DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature* **421**, 499–506 (2003).
34. Lee, J. H. & Paull, T. T. ATM activation by DNA double-strand breaks through the Mre11–Rad50–Nbs1 complex. *Science* **308**, 551–554 (2005).
35. Lee, J. H. & Paull, T. T. Direct activation of the ATM protein kinase by the Mre11/Rad50/Nbs1 complex. *Science* **304**, 93–96 (2004).
36. Francke, U. Williams–Beuren syndrome: genes and mechanisms. *Hum. Mol. Genet.* **8**, 1947–1954 (1999).
37. Pear, W., Scott, M. & Nolan, G. P. in *Methods in Molecular Medicine: Gene Therapy Protocols*, 41–57 (Humana, 1997).
38. Bellard, M., Dretzen, G., Giangrande, A. & Romain, P. Nuclease digestion of transcriptionally active chromatin. *Methods Enzymol.* **170**, 317–346 (1989).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank A. Nussenzweig for H2A.X<sup>-/-</sup> MEF cells, P. Varga-Weisz for WSTF constructs and antibodies, W. Herr for pBABE-puro vectors, J. Kim and R. Roeder for anti-BAF53 antibodies, D. Reinberg for anti-SNF2H antibodies, Z. Lou for anti-Mdc1 antibodies and Mdc1<sup>-/-</sup> MEF cells, L. Liang and Q. Li for their assistance in recombinant protein expression and purification, C. H. McDonald, R. G. Cook and The Rockefeller University Proteomic Core facility for H2A.X peptides. We would also like to thank the Millipore antibody development scientists for collaborating with us on the generation of H2A.X Tyr142(ph) antibodies, catalogue number 07-1590. This study was supported by the following sources: Susan G. Komen Breast Cancer Foundation (A.X.), Abby Rockefeller Mauze Trust and Starr Foundation (H.L. and D.J.P.), The Dewitt Wallace and Maloris Foundations (H.L. and D.J.P.), The Irma T. Hirsch Trust (D.S.), NCI Cancer Center Support Grant P30 CA08748 (L.A.F., H.E.-B. and P.T.), research grants from National Institutes of Health to S.J.E., S.H.A. and C.D.A., and The Rockefeller University (C.D.A.). S.J.E. is an Investigator with the Howard Hughes Medical Institute. We are grateful to E. Bernstein and E. Duncan for critical reading of the manuscript.

**Author Contributions** A.X. designed the study, performed the experiments and wrote the paper; H.L. generated recombinant WSTF protein and performed CD analysis; D.S. helped with the experiments performed in *Xenopus* egg extracts and edited the manuscript; S.H.A. generated and analysed H2A Leu132Tyr mutant yeast strain; L.A.F., H.E.-B. and P.T. performed MS analysis; S.I.-M. provided technical assistance for protein production; B.W. and S.J.E. provided Mdc1 constructs; K.H. performed bioinformatical analysis on the WSTF kinase domain; D.J.P. provided general guidance for generating recombinant WSTF. S.J.E. also discussed results and commented on the manuscript. C.D.A. provided support and general guidance for this work.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to C.D.A. ([alliscd@rockefeller.edu](mailto:alliscd@rockefeller.edu)).



## METHODS

**Plasmids and cell culture.** Tyr 142 mutant constructs were derived from wild-type human H2A.X plasmids (Openbiosystems) using Quickchange kit (Stratagene) and verified by DNA sequencing. The primers were: Tyr142Leu forward, 5'-CAAGAAGGCCACCCAGGCTCCAGGAGCTCTAAG-3'; Tyr142Leu reverse, 5'-CTTAGAGCTCCTGGGAGGCTGGGTGGCCTTCTTG-3'; Tyr142Phe forward, 5'-CAAGAAGGCCACCCAGGCTCCAGGAGTCTAAG-3'; Tyr142Leu reverse, 5'-CTTAGAAGCTCCTGGGAGGCTGGGTGGCCTTCTTG-3'. These plasmids were amplified by PCR and cloned into pTOPO vectors (Invitrogen). The primers were: wild-type forward, 5'-CACCTCGGGCCGCGCAAGA-CTG-3'; wild-type reverse, 5'-CTTAGTACTCCTGGGAG-3'; Tyr142Leu, 5'-CTTAGAGCTCCTGGGAG-3'; Tyr142Phe, 5'-CTTAGAAGCTCCTGGGAG-3'.

To generate N-terminal Flag-tagged constructs, the wild-type or mutant H2A.X coding region in the pTOPO vectors were cloned in-frame between the BamHI and XhoI sites of the pCMV-Tag2A vectors (Stratagene). Tagged or untagged constructs were cloned into MSCV-puro vectors (gifts from W. Herr's laboratory). To generate constructs for recombinant protein expression in *E. coli*, the coding regions of the wild-type or Tyr142Phe H2A.X of the untagged vectors (above) were cloned into pET28 vectors (Novagen), in-frame with the N-terminal 6×His tag.

MSCV virus production and infection followed standard protocols<sup>37</sup>. In brief, MSCV viruses were packaged in Phoenix cells and used to infect H2A.X<sup>-/-</sup> MEFs (gifts from A. Nussenzweig's laboratory), followed by puromycin selection. For each construct, three independent lines were derived. The expression levels of H2A.X were checked by immunoblots and immunofluorescence.

Full-length WSTF (human) constructs was a gift from P. Varga-Weisz. To generate a series of truncated WSTF constructs for insect cell expression, the PCR fragments of the corresponding regions of the WSTF gene were cloned into a modified pFASTBac vector (Invitrogen), which contains an N-terminal GST and a C-terminal His-tag sequence. The Cys338Ala mutant (1–359) was derived from the wild-type 1–359 construct, using the QuickChange site-directed mutagenesis kit (Stratagene) and was verified by DNA sequencing. To generate constructs for protein expression in *E. coli*, the PCR product of the N-motif (1–205) was cloned into pGex6p vector (GE Healthcare) in-frame with the N-terminal GST tag, and the PCR product of the C-motif (208–345) was cloned into a modified pRSFDuet-1 vector (Novagen) in-frame with the N-terminal MBP (maltose binding protein) tag. To generate constructs for mammalian cell expression, the PCR products (the forward and reverse primers contains a BamHI and XhoI site sequence, respectively) of the corresponding fragments of the human WSTF gene were cloned in-between the BamHI and XhoI sites of pTAG3A vectors (Stratagene), in-frame with the N-terminal Myc epitope tag. The tagged constructs were cloned into MSCV-puro vectors.

**Generation of WSTF RNAi knockdown cells.** WSTF RNAi cell lines were generated using recombinant pShag-2 vectors containing short hairpin RNA (shRNA) constructs targeting the mouse *Wstf* gene (Clone ID: V2MM\_17087, Openbiosystems; hairpin sequence: GCTGTTGACAGTGAGCGCGGAGATAC-TTCGATACTTTATTAGTGAAGCCACAGATGTAATAAAGTATCGAAGTATCTCTTGGCTTGCCTCGGA). Three independent lines were generated by infecting virus carrying these vectors into NIH3T3 cells and propagated under puromycin. In Fig. 4, the WSTF RNAi cells were infected with recombinant MSCV viruses carrying the wild-type (1–359) or the mutant (Cys338Ala) constructs.

**Purification of H2A.X-containing mononucleosomes and associated protein factors.** MEF fractionation and chromatin pellet isolation (Fig. 2) were performed as described<sup>22</sup>. Chromatin pellets were briefly digested with Mnase (Sigma) and the generation of mononucleosomes was monitored by electrophoresis<sup>38</sup>. Fifty microlitres of Flag-M2 agarose beads were used for immunoprecipitating the total mononucleosomes isolated from  $2 \times 10^8$  cells. The precipitated complex was resolved on 4–12% gradient gels (Invitrogen) and then silver or Coomassie-blue stained. The gel bands were isolated and subjected to mass spectrometry.

**Protein identification by mass spectrometry.** Gel-resolved protein were digested with trypsin, batch purified on a reversed-phase micro-tip, and the resulting peptide pools individually analysed by matrix-assisted laser desorption/ionization reflectron time-of-flight (MALDI-reTOF) mass spectrometry (UltraFlex TOF/TOF; Bruker Daltonics) for peptide mass fingerprinting (PMF), as described<sup>39,40</sup>. Selected peptide ions (*m/z*) were taken to search a 'non-redundant' protein database ('NR'; 3,245,378 entries on 28 January 2006; National Center for Biotechnology Information) using the PeptideSearch algorithm (M. Mann, Max-Planck Institute for Biochemistry; an updated version of this program is at present available as 'PepSea' from Applied Biosystems/MDS Sciex). A molecular mass range up to twice the apparent molecular weight (as estimated from electrophoretic relative mobility) was covered, with a mass accuracy restriction of less than 40 p.p.m., and a maximum of one missed cleavage site allowed per peptide. To confirm the PMF results, mass spectrometric sequencing

of selected peptides was performed by MALDI-TOF/TOF (MS/MS) analysis on the same prepared samples, using the UltraFlex instrument in 'LIFT' mode. Fragment ion spectra were taken to search the NR database using the MASCOT MS/MS Ion Search program<sup>41</sup>, version 2.0.04 for Windows (Matrix Science Ltd.). Any tentative confirmation of a PMF result thus obtained was verified by comparing the computer-generated fragment ion series of the predicted tryptic peptide with the experimental MS/MS data.

**Immunoblotting and immunofluorescence.** Cells at 70–80% confluence were exposed to an ionizing radiation source (Cs) to introduce DNA damage (at 10 Gy). Samples were collected at the indicated time points after irradiation. Histone extraction, cell fractionation, high salt nuclear extraction and western blotting followed standard protocols<sup>42,43</sup>. For immunofluorescence, cells were grown on coverslips and fixed with 3% paraformaldehyde. The subsequent indirect immunofluorescent staining followed standard protocols<sup>43</sup>. For WSTF staining, cells were extracted with a high-salt buffer for 2 min at 4 °C before fixing, as described<sup>28</sup>. The antibodies used were:  $\gamma$ -H2A.X, anti-general H2A.X and anti-BRCA1 (Millipore/Upstate); anti-WSTF and anti-Myc (Sigma); anti-ATM Ser 1981(ph) (Rockland immunological); anti-Mdc1, anti-BAF53 and anti-SNF2H were gifts from Z. Lou, J. Kim and R. Roeder, and D. Reinberg, respectively.

**Recombinant Mdc1 protein expression and peptide pull-down assay.** GST-Mdc1-BRCT protein expression followed manufacturer's protocols (Stratagene). Peptides (UN (unmodified): N-CPSGGKKATQASQEY-C; Ser 139(ph): N-CPSGGKKATQApSQEY-C; Tyr 142(ph): N-CPSGGKKATQASQEpY-C; Ser 139(ph) and Tyr 142(ph): N-CPSGGKKATQApSQEY-C; Phe 142: N-CPSGGKKATQASQEF-C; Leu 142: N-CPSGGKKATQASQEL-C S139(ph), Leu 142: N-CPSGGKKATQApSQEL-C; Ser 139(ph), Phe 142: N-CPSGGKKATQApSQEF-C) were conjugated to agarose bead with Sulfolink kit (Pierce). Peptide pull-down experiments were performed as described previously<sup>44</sup>.

**Circular dichroism measurement.** Circular dichroism (CD) experiments were performed on an Aviv 62A/DS (Aviv Associates) spectropolarimeter. Spectra (260 nm to 200 nm) were collected with a cell-path length of 0.1 cm and a bandwidth of 1 nm. Samples were diluted to 10  $\mu$ M with 20 mM Tris-HCl, pH 7.5, 100 mM KCl, 2 mM dithiothreitol. Protein concentrations were determined with their 280 nm absorbance.

**Automated chemical ('Edman') protein sequencing.** Protein samples were analysed with a Procise 494 instrument from Applied Biosystems as described<sup>45</sup>. Stepwise liberated phenylthiohydantoin (PTH)-amino acids were identified using an 'online' HPLC system (Applied Biosystems) equipped with a PTH C18 (2.1  $\times$  220 mm; 5 micrometre particle size) column (Applied Biosystems).

**Mapping WSTF kinase domain and preparation of recombinant WSTF protein from insect cells.** To map the putative kinase domain, we generated a series of protein constructs within the N-terminal regions of WSTF. All these protein constructs were extensively purified (Fig. 3a, right), and mass spectrometry analysis did not identify any contaminating protein kinases. The 1–345 construct had more potent kinase activity (>50-fold) than the 1–340 construct (Fig. 3e). Furthermore, the substrate specificity of the 1–345 construct is similar to the full-length WSTF (Supplementary Fig. 5). These results indicate that the amino acid residues in the surrounding regions of 340–345 are critical for the kinase activity of WSTF. Consistent with these results, the activity of a point mutant at Cys 338 (Cys338Ala) was greatly diminished (<50-fold, Fig. 3f). The CD spectrum of this mutant is identical to the wild-type counterpart (Supplementary Fig. 6), indicating that this mutation did not alter the global structure of the protein.

Recombinant baculovirus carrying the full-length WSTF (Flag-epitope-tagged at C terminus) was produced using the BaculoGold-BEVS kit (BD Biosciences). The full-length WSTF proteins were purified using a published protocol<sup>23</sup>. The viruses that express truncated WSTF constructs were produced using the Bac-to-Bac (Invitrogen) method, all of which were tagged with an N-terminal GST tag and a C-terminal 6×His tag. Baculoviruses were produced in Sf9 cells, and proteins were expressed in Hi5 suspension cells. For truncated WSTF protein purification, collected Hi5 cells were suspended in 0.4 M KCl, 20 mM Tris, pH 7.5, supplemented with the EDTA-free Complete Protease Inhibitor Cocktail tablet (Roche). Whole cell lysates were purified through GSTrap columns (GE Healthcare). After removal of the GST tags by PreScission Protease digestion (GE Healthcare), the protein mixtures were purified by HiTrap SP (GE Healthcare). The recombinant WSTF proteins (peaks form at ~200 mM KCl, 20 mM Tris, pH 7.5) were further purified with Superdex 75 (GE Healthcare) in the elution buffer (50 mM KCl, 20 mM Tris, pH 7.5). The monomeric peaks (~60 kDa) were collected and concentrated for further use. The Cys338Ala mutant protein was expressed and purified as the wild-type protein described above.

**Protein preparation in *E. coli*.** The recombinant plasmid vectors carrying the N- and C-motif constructs (described earlier) were transformed into *E. coli* host

strain Rosetta2 (DE3) (Novagen) for protein expression under triple antibiotic selection (ampicillin, kanamycin and chloramphenicol) in LB medium. After overnight induction by 0.4 mM isopropyl  $\beta$ -D-thiogalactoside at 25 °C, cells were collected in buffer (0.4 M KCl, 20 mM Tris, pH 7.5). Co-expressed proteins were affinity purified using Amylose resin (New England Biolabs) against the MBP tag. Maltose-eluted samples were purified through a Superdex 200 column (GE Healthcare) in the elution buffer (50 mM KCl, 20 mM Tris, pH 7.5). Most proteins were eluted in a single peak that contains GST-tagged WAC N-motif and MBP-tagged WAC C-motif with a stoichiometry of 1:1.

The N- or C-motif were also expressed separately in Rosetta2 (DE3) (Novagen) cells in parallel. The N-motif was purified with GSTrap column and Superdex 200 and the C-motif was purified with Amylose resin and Superdex 200. Most proteins were eluted in a monomeric peak (~50 kDa).

**Histone protein purification and *in vitro* assembly of histone octamers and mononucleosomes.** Individual free histone proteins (H2A.X, H2B, H3 and H4) were expressed and purified using a standard protocol<sup>46</sup>. N-terminal 6 $\times$ His-epitope-tagged wild-type and Phe 142 H2A.X proteins were purified with Ni Sepharose6 columns (GE Healthcare). The assembly of the histone octamers followed standard protocols<sup>46</sup>. Four recombinant histone proteins were denatured in 8 M guanidine-HCl solutions and dialysed to 2 M NaCl. The crude histone octamers were further purified through a Superdex 75 column (GE Healthcare). Mononucleosomes were assembled as described<sup>47</sup>. The PCR products of the DNA template (nucleosome positioning sequence (clone 601)<sup>48</sup>) was initially incubated in a reaction buffer<sup>47</sup> supplemented with 2 M NaCl and diluted stepwise in the same reaction buffer until the NaCl concentration reached 200 mM. The assembly of mononucleosomes was checked by electrophoresis on 5% native PAGE gels.

***In vitro* kinase assays.** Recombinant full-length (50–100 ng) or truncated WSTF proteins (0.1–0.5  $\mu$ g) were incubated with free histone proteins (1–2  $\mu$ g), histone octamers or nucleosomes assembled *in vitro* (0.5–1  $\mu$ g) in 20 mM Tris, pH 7.4 and 150 mM NaCl supplemented with  $\gamma$ -<sup>32</sup>P-ATP (NEN) and 1 mM MnCl<sub>2</sub> for 45 min. In Supplementary Fig. 4, 0.1–1 mM of MnCl<sub>2</sub> or MgCl<sub>2</sub> was supplemented as indicated. The reaction was stopped by the addition of 2 mM EDTA. The reaction mixtures were separated by SDS-PAGE, stained with Coomassie blue and dried. The dried gels were exposed against phosphorimager screens (FujiFilm). The autoradiography images were scanned in phosphorimager (FujiFilm) and analysed by Image Reader FLA-5000.

39. Erdjument-Bromage, H. *et al.* Examination of micro-tip reversed-phase liquid chromatographic extraction of peptide pools for mass spectrometric analysis. *J. Chromatogr. A* **826**, 167–181 (1998).
40. Sebastiaan Winkler, G. *et al.* Isolation and mass spectrometry of transcription factor complexes. *Methods* **26**, 260–269 (2002).
41. Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567 (1999).
42. Lennox, R. W. & Cohen, L. H. Analysis of histone subtypes and their modified forms by polyacrylamide gel electrophoresis. *Methods Enzymol.* **170**, 532–549 (1989).
43. Sambrook, J. & Russell, D. W. *Molecular Cloning: a Laboratory Manual* (Cold Spring Harbour Laboratory Press, 2001).
44. Wysocka, J. *et al.* A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**, 86–90 (2006).
45. Tempst, P., Geromanos, S., Elicone, C. & Erdjument-Bromage, H. Improvements in microsequencer performance for low picomole sequence analysis. *Methods* **6**, 248–261 (1994).
46. Luger, K., Rechsteiner, T. J. & Richmond, T. J. Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol.* **304**, 3–19 (1999).
47. Steger, D. J., Eberharter, A., John, S., Grant, P. A. & Workman, J. L. Purified histone acetyltransferase complexes stimulate HIV-1 transcription from preassembled nucleosomal arrays. *Proc. Natl Acad. Sci. USA* **95**, 12924–12929 (1998).
48. Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19–42 (1998).



# A role for self-gravity at multiple length scales in the process of star formation

Alyssa A. Goodman<sup>1,2</sup>, Erik W. Rosolowsky<sup>2,3</sup>, Michelle A. Borkin<sup>1†</sup>, Jonathan B. Foster<sup>2</sup>, Michael Halle<sup>1,4</sup>, Jens Kauffmann<sup>1,2</sup> & Jaime E. Pineda<sup>2</sup>

Self-gravity plays a decisive role in the final stages of star formation, where dense cores (size  $\sim 0.1$  parsecs) inside molecular clouds collapse to form star-plus-disk systems<sup>1</sup>. But self-gravity's role at earlier times (and on larger length scales, such as  $\sim 1$  parsec) is unclear; some molecular cloud simulations that do not include self-gravity suggest that 'turbulent fragmentation' alone is sufficient to create a mass distribution of dense cores that resembles, and sets, the stellar initial mass function<sup>2</sup>. Here we report a 'dendrogram' (hierarchical tree-diagram) analysis that reveals that self-gravity plays a significant role over the full range of possible scales traced by  $^{13}\text{CO}$  observations in the L1448 molecular cloud, but not everywhere in the observed region. In particular, more than 90 per cent of the compact 'pre-stellar cores' traced by peaks of dust emission<sup>3</sup> are projected on the sky within one of the dendrogram's self-gravitating 'leaves'. As these peaks mark the locations of already-forming stars, or of those probably about to form, a self-gravitating cocoon seems a critical condition for their existence. Turbulent fragmentation simulations without self-gravity—even of unmagnetized isothermal material—can yield mass and velocity power spectra very similar to what is observed in clouds like L1448. But a dendrogram of such a simulation<sup>4</sup> shows that nearly all the gas in it (much more than in the observations) appears to be self-gravitating. A potentially significant role for gravity in 'non-self-gravitating' simulations suggests inconsistency in simulation assumptions and output, and that it is necessary to include self-gravity in any realistic simulation of the star-formation process on subparsec scales.

Spectral-line mapping shows whole molecular clouds (typically tens to hundreds of parsecs across, and surrounded by atomic gas) to be marginally self-gravitating<sup>5</sup>. When attempts are made to further break down clouds into pieces using 'segmentation' routines, some self-gravitating structures are always found on whatever scale is sampled<sup>6,7</sup>. But no observational study to date has successfully used one spectral-line data cube to study how the role of self-gravity varies as a function of scale and conditions, within an individual region.

Most past structure identification in molecular clouds has been explicitly non-hierarchical, which makes difficult the quantification of physical conditions on multiple scales using a single data set. Consider, for example, the often-used algorithm CLUMPFIND<sup>7</sup>. In three-dimensional (3D) spectral-line data cubes, CLUMPFIND operates as a watershed segmentation algorithm, identifying local maxima in the position–position–velocity ( $p$ – $p$ – $v$ ) cube and assigning nearby emission to each local maximum. Figure 1 gives a two-dimensional (2D) view of L1448, our sample star-forming region, and Fig. 2 includes a CLUMPFIND decomposition of it based on  $^{13}\text{CO}$  observations. As with any algorithm that does not offer hierarchically nested or

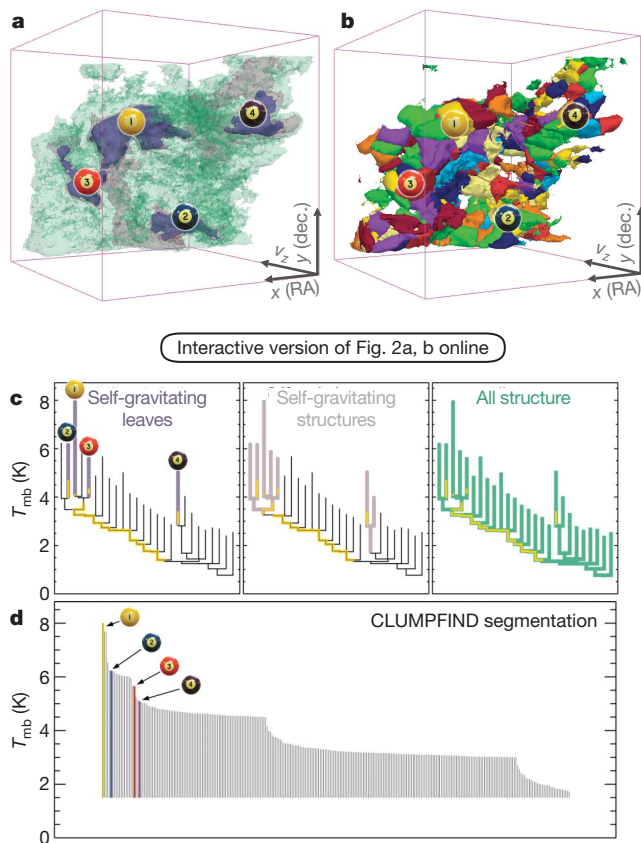
overlapping features as an option, significant emission found between prominent clumps is typically either appended to the nearest clump or turned into a small, usually 'pathological', feature needed to encompass all the emission being modelled. When applied to molecular-line



**Figure 1 | Near-infrared image of the L1448 star-forming region with contours of molecular emission overlaid.** The channels of the colour image correspond to the near-infrared bands *J* (blue), *H* (green) and *K* (red), and the contours of integrated intensity are from  $^{13}\text{CO}(1-0)$  emission<sup>8</sup>. Integrated intensity is monotonically, but not quite linearly (see Supplementary Information), related to column density<sup>18</sup>, and it gives a view of 'all' of the molecular gas along lines of sight, regardless of distance or velocity. The region within the yellow box immediately surrounding the protostars has been imaged more deeply in the near-infrared (using Calar Alto) than the remainder of the box (2MASS data only), revealing protostars as well as the scattered starlight known as 'Cloudshine'<sup>21</sup> and outflows (which appear orange in this colour scheme). The four billiard-ball labels indicate regions containing self-gravitating dense gas, as identified by the dendrogram analysis, and the leaves they identify are best shown in Fig. 2a. Asterisks show the locations of the four most prominent embedded young stars or compact stellar systems in the region (see Supplementary Table 1), and yellow circles show the millimetre-dust emission peaks identified as star-forming or 'pre-stellar' cores<sup>3</sup>.

<sup>1</sup>Initiative in Innovative Computing at Harvard, Cambridge, Massachusetts 02138, USA. <sup>2</sup>Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA.

<sup>3</sup>Department of Physics, University of British Columbia, Okanagan, Kelowna, British Columbia V1V 1V7, Canada. <sup>4</sup>Surgical Planning Laboratory and Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>†</sup>Present address: School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA.



**Figure 2 | Comparison of the 'dendrogram' and 'CLUMPFIND' feature-identification algorithms as applied to  $^{13}\text{CO}$  emission from the L1448 region of Perseus.** **a**, 3D visualization of the surfaces indicated by colours in the dendrogram shown in **c**. Purple illustrates the smallest scale self-gravitating structures in the region corresponding to the leaves of the dendrogram; pink shows the smallest surfaces that contain distinct self-gravitating leaves within them; and green corresponds to the surface in the data cube containing all the significant emission. Dendrogram branches corresponding to self-gravitating objects have been highlighted in yellow over the range of  $T_{mb}$  (main-beam temperature) test-level values for which the virial parameter is less than 2. The  $x$ - $y$  locations of the four 'self-gravitating' leaves labelled with billiard balls are the same as those shown in Fig. 1. The 3D visualizations show position–position–velocity ( $p$ - $p$ - $v$ ) space. RA, right ascension; dec., declination. For comparison with the ability of dendrograms (**c**) to track hierarchical structure, **d** shows a pseudo-dendrogram of the CLUMPFIND segmentation (**b**), with the same four labels used in Fig. 1 and in **a**. As 'clumps' are not allowed to belong to larger structures, each pseudo-branch in **d** is simply a series of lines connecting the maximum emission value in each clump to the threshold value. A very large number of clumps appears in **b** because of the sensitivity of CLUMPFIND to noise and small-scale structure in the data. In the online PDF version, the 3D cubes (**a** and **b**) can be rotated to any orientation, and surfaces can be turned on and off (interaction requires Adobe Acrobat version 7.0.8 or higher). In the printed version, the front face of each 3D cube (the 'home' view in the interactive online version) corresponds exactly to the patch of sky shown in Fig. 1, and velocity with respect to the Local Standard of Rest increases from front ( $-0.5 \text{ km s}^{-1}$ ) to back ( $8 \text{ km s}^{-1}$ ).

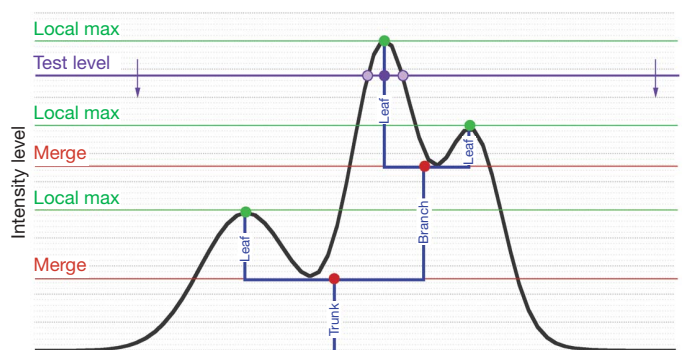
data, CLUMPFIND typically finds features on a limited range of scales, above but close to the physical resolution of the data, and its results can be overly dependent on input parameters. By tuning CLUMPFIND's two free parameters, the same molecular-line data set<sup>8</sup> can be used to show either that the frequency distribution of clump mass is the same as the initial mass function of stars or that it follows the much shallower mass function associated with large-scale molecular clouds (Supplementary Fig. 1).

Four years before the advent of CLUMPFIND, 'structure trees'<sup>9</sup> were proposed as a way to characterize clouds' hierarchical structure

using 2D maps of column density. With this early 2D work as inspiration, we have developed a structure-identification algorithm that abstracts the hierarchical structure of a 3D ( $p$ - $p$ - $v$ ) data cube into an easily visualized representation called a 'dendrogram'<sup>10</sup>. Although well developed in other data-intensive fields<sup>11,12</sup>, it is curious that the application of tree methodologies so far in astrophysics has been rare, and almost exclusively within the area of galaxy evolution, where 'merger trees' are being used with increasing frequency<sup>13</sup>.

Figure 3 and its legend explain the construction of dendrograms schematically. The dendrogram quantifies how and where local maxima of emission merge with each other, and its implementation is explained in Supplementary Methods. Critically, the dendrogram is determined almost entirely by the data itself, and it has negligible sensitivity to algorithm parameters. To make graphical presentation possible on paper and 2D screens, we 'flatten' the dendrograms of 3D data (see Fig. 3 and its legend), by sorting their 'branches' to not cross, which eliminates dimensional information on the  $x$  axis while preserving all information about connectivity and hierarchy. Numbered 'billiard ball' labels in the figures let the reader match features between a 2D map (Fig. 1), an interactive 3D map (Fig. 2a online) and a sorted dendrogram (Fig. 2c).

A dendrogram of a spectral-line data cube allows for the estimation of key physical properties associated with volumes bounded by iso-surfaces, such as radius ( $R$ ), velocity dispersion ( $\sigma_v$ ) and luminosity ( $L$ ). The volumes can have any shape, and in other work<sup>14</sup> we focus on the significance of the especially elongated features seen in L1448 (Fig. 2a). The luminosity is an approximate proxy for mass, such that  $M_{\text{lum}} = X_{^{13}\text{CO}} L_{^{13}\text{CO}}$ , where  $X_{^{13}\text{CO}} = 8.0 \times 10^{20} \text{ cm}^2 \text{ K}^{-1} \text{ km}^{-1} \text{ s}$  (ref. 15; see Supplementary Methods and Supplementary Fig. 2). The derived values for size, mass and velocity dispersion can then be used to estimate the role of self-gravity at each point in the hierarchy, via calculation of an 'observed' virial parameter,  $\alpha_{\text{obs}} = 5\sigma_v^2 R / GM_{\text{lum}}$ . In principle, extended portions of the tree (Fig. 2, yellow highlighting) where  $\alpha_{\text{obs}} < 2$  (where gravitational energy is comparable to or larger than kinetic energy) correspond to regions of  $p$ - $p$ - $v$  space where self-gravity is significant. As  $\alpha_{\text{obs}}$  only represents the ratio of kinetic energy to gravitational energy at one point in time, and does not explicitly capture external over-pressure and/or magnetic fields<sup>16</sup>, its measured value should only be used as a guide to the longevity (boundedness) of any particular feature.



**Figure 3 | Schematic illustration of the dendrogram process.** Shown is the construction of a dendrogram from a hypothetical one-dimensional emission profile (black). The dendrogram (blue) can be constructed by 'dropping' a test constant emission level (purple) from above in tiny steps (exaggerated in size here, light lines) until all the local maxima and mergers are found, and connected as shown. The intersection of a test level with the emission is a set of points (for example the light purple dots) in one dimension, a planar curve in two dimensions, and an isosurface in three dimensions. The dendrogram of 3D data shown in Fig. 2c is the direct analogue of the tree shown here, only constructed from 'isosurface' rather than 'point' intersections. It has been sorted and flattened for representation on a flat page, as fully representing dendrograms for 3D data cubes would require four dimensions.

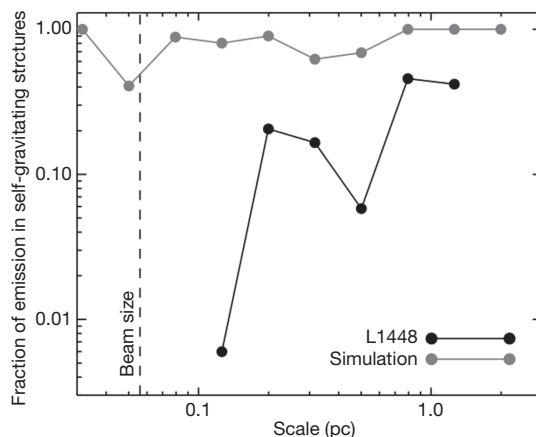


In calculating  $\alpha_{\text{obs}}$ , we are implicitly assuming that there is a one-to-one relationship (known as a ‘bijection’) between a volume in  $p$ - $p$ - $v$  space and a volume of physical (position–position–position,  $p$ - $p$ - $p$ ) space. This bijection paradigm is fine for regions which are dominated by a single structure, but the complexities of relating  $p$ - $p$ - $v$  space to physical space in regions with multiple features along a line of sight does mean that this treatment can only ever give an approximate measure of the true dynamical state of the cloud<sup>17</sup>. Alternatives to bijection are considered in the Supplementary Information. The bijection assumption comes into play when measuring physical properties of individual features, but it does not influence the characterization of hierarchical structure.

In Fig. 2c, we show the dendrogram for the same L1448 <sup>13</sup>CO spectral-line map shown using contours in Fig. 1. All of the portions shaded yellow have  $\alpha_{\text{obs}} < 2$ , meaning that they are (most) likely to be self-gravitating. The four most compact  $p$ - $p$ - $v$  structures (leaves) where  $\alpha_{\text{obs}} < 2$  are numbered in Figs 1 and 2, and they are not as apparent in the projected (2D) view (Fig. 1) as they are in  $p$ - $p$ - $v$  (3D) space (Fig. 2a). In the CLUMPFIND decomposition of the cloud (Fig. 2b), these features are not apparent as special.

Overall, the pattern of yellow highlighting in Fig. 2 suggests the importance of gravity on all possible scales, but not within the full possible volume, in a cloud like L1448. With the exception of the gas around region 4, which appears not to be bound to the rest of L1448, the tree shows a fully yellow-highlighted ‘trunk’ and only sporadic highlighting on the dendrogram’s tallest branches and leaves. So for the material traced by <sup>13</sup>CO observations, it appears that self-gravitating structures are more prevalent on larger scales than on smaller. At densities surpassing  $5 \times 10^3 \text{ cm}^{-3}$ , <sup>13</sup>CO becomes an increasingly poor tracer of mass<sup>18</sup>, so it can only give upper limits for the ‘true’ virial parameters of the densest, most compact, structures seen in the dendrogram. Thus, the highest-density non-yellow leaves in Fig. 2c may harbour bound structures only visible with thinner or less-depleted molecular lines. On the other hand, lower-density non-yellow leaves in Fig. 2c probably represent actual low-mass unbound structures in the gas, similar to the ‘pressure-confined’ low-mass clumps found in clump-based segmentations. Importantly, the full pattern of highlighting explicitly indicates that core-like leaves often reside within structures where the mutual gravity between the cores (leaves) and/or their environs (branches) is significant enough to cause meaningful interactions between cores—possibly even, in the most extreme cases, competitive accretion. Recent work<sup>18</sup> has shown that the overall (column) density distribution of material traced by <sup>13</sup>CO in a 10-pc-scale molecular cloud is roughly log-normal, and our result here implies that some of the high-density fluctuations in that statistical distribution are bound within themselves and/or to each other, and some not.

Tree hierarchies can be used to intercompare the topology and physical properties (for example boundedness) of structures within star-forming regions, and such intercomparison can be profitably extended to simulations as well. In Fig. 4, we summarize such a comparison (see Supplementary Information) with a plot showing the fraction of ‘self-gravitating’ ( $\alpha_{\text{obs}} < 2$ ) material as a function of spatial scale for both our L1448 data and for a synthetic data cube<sup>4</sup>. The simulation used to produce the synthetic data is purely hydrodynamic, meaning that the effects of magnetic fields, heating and cooling, and self-gravity are not included. The power-law exponent characterizing the power spectrum of turbulence in these synthetic <sup>13</sup>CO data and in the COMPLETE Perseus data<sup>8</sup> (from which our L1448 example is drawn) is  $\sim 1.8$ , to within small uncertainties ( $\sim 0.2$ ; ref. 4). However, inspection of Fig. 4 (and of Supplementary Fig. 4) clearly shows that the data and simulation appear quite different in the context of dendrogram analysis: in the simulation, nearly all material (much more than in the observations) is self-gravitating, on all spatial scales. Critically, the analysis of the synthetic <sup>13</sup>CO cube<sup>4</sup> (Supplementary Fig. 4) is done on a simulated observation of it where we have deliberately matched resolution,



**Figure 4 | The fraction of self-gravitating emission as a function of scale in L1448 and a comparable simulation.** Most of the emission in the L1448 region is contained within large-scale self-gravitating structures, but only a low fraction of small-scale objects show signs of self-gravitation. (See text for discussion of the high-density, small-scale, self-gravitating structures to which <sup>13</sup>CO is insensitive.) In the L1448 observations, gravity is significant on all scales, but not in all regions. In contrast, the simulated map implies that nearly all scales, and all regions, should be influenced by gravity.

noise properties and region extent to the L1448 cube (Supplementary Methods). The (constant) abundance of <sup>13</sup>CO used for the synthetic map (Supplementary Information) is set to match the known column densities in the simulation, and because abundance is simply a multiplicative constant, changing it cannot reproduce the scale dependence of gravity found in the L1448 data.

Thus it appears that the synthetic data cube created from the simulation<sup>4</sup> contains much material that would be significantly affected by gravity, if gravity were actually included in the simulation.

The accuracy with which dendrograms can offer estimates of  $\alpha_{\text{obs}}$  is at or below the 25% level (Supplementary Information). The uncertainty results primarily from the need to glean a 3D geometry and density based on 2D size and column density (mass/area), and any analysis of  $p$ - $p$ - $v$  data will be subject to the same limitations. More analysis, using simulations, of the translation from  $p$ - $p$ - $v$  to  $p$ - $p$ - $p$  space<sup>17</sup> should be, and is being, carried out to quantify these uncertainties more finely. Comparative measurements (for example Fig. 4) are far more certain as these biases should affect all data sets similarly. Thus, the apparent disagreement between observations and simulation in Fig. 4 can be explained by claiming that either, or both, of the following are true: (1) the assumptions/calculations leading to the creation of the synthetic <sup>13</sup>CO observations are faulty; or (2) there is missing physics in the simulation (for example gravity, thermal effects), making it an insufficient approximation to real star-forming regions.

Finally, we turn to the relationship between the apparently ‘self-gravitating’ regions in L1448 and the star-formation process itself. Compact millimetre-wavelength emission peaks caused by dust emission (marked by yellow circles in Fig. 1) are typically taken as markers of cores that are forming, or are able to form, stars. Within the region of L1448 considered here, more than 90% of the compact millimetre-dust peaks traced in bolometer observations<sup>3</sup> are found projected on the sky within one of the dendrogram’s ‘self-gravitating’ leaves, and none is found outside a self-gravitating branch. Recent NH<sub>3</sub> observations<sup>19</sup> suggest that all, or all but one, of these ‘pre-stellar cores’ lie within self-gravitating structures along the velocity dimension as well<sup>14</sup>. As young sources get a little older, they can be detected in the mid-infrared (IRAC) bands of the Spitzer Space Telescope. Four out of the five sources identified by such IRAC imaging as protostar candidates<sup>20</sup> also lie within a leaf, and each of those four is associated with a millimetre-dust peak, suggesting they are embedded in dense natal cocoons. Interestingly, the one IRAC protostar

candidate in the region not associated with a self-gravitating leaf is also not associated with a millimetre-dust peak, suggesting it is a more evolved source. All told, these associations suggest that a self-gravitating home is critical to the earliest phases of star formation.

Received 28 June 2007; accepted 28 October 2008.

- Di Francesco, J. *et al.* in *Protostars and Planets V* (eds Reipurth, B., Jewitt, D. & Keil, K.) 17–32 (Univ. Arizona Press, 2006).
- Padoan, P. & Nordlund, Å. The stellar initial mass function from turbulent fragmentation. *Astrophys. J.* **576**, 870–879 (2002).
- Enoch, M. L. *et al.* Bolocam survey for 1.1 mm dust continuum emission in the c2d legacy clouds. I. Perseus. *Astrophys. J.* **638**, 293–313 (2006).
- Padoan, P., Juvela, M., Kritsuk, A. & Norman, M. L. The power spectrum of supersonic turbulence in Perseus. *Astrophys. J.* **653**, L125–L128 (2006).
- Larson, R. B. Turbulence and star formation in molecular clouds. *Mon. Not. R. Astron. Soc.* **194**, 809–826 (1981).
- Stutzki, J. & Gusten, R. High spatial resolution isotopic CO and CS observations of M17 SW: The clumpy structure of the molecular cloud core. *Astrophys. J.* **356**, 513–515 (1990).
- Williams, J., de Geus, E. & Blitz, L. Determining structure in molecular clouds. *Astrophys. J.* **428**, 693–712 (1994).
- Ridge, N. A. *et al.* The COMPLETE survey of star-forming regions: Phase I data. *Astron. J.* **131**, 2921–2933 (2006).
- Houllahan, P. & Scalo, J. Recognition and characterization of hierarchical interstellar structure. II - Structure tree statistics. *Astrophys. J.* **393**, 172–187 (1992).
- Rosolowsky, E. W., Pineda, J. E., Kauffmann, J. & Goodman, A. A. Structural analysis of molecular clouds: Dendrograms. *Astrophys. J.* **679**, 1338–1351 (2008).
- Heine, C., Scheuermann, G., Flamm, C., Hofacker, I. L. & Stadler, P. F. Visualization of barrier tree sequences. *IEEE Trans. Vis. Comput. Graph.* **12**, 781–788 (2006).
- Vliegen, R., van Wijk, J. J. & van der Linden, E.-J. Visualizing business data with generalized treemaps. *IEEE Trans. Vis. Comput. Graph.* **12**, 789–796 (2006).
- Kauffmann, G. & White, S. D. M. The merging history of dark matter haloes in a hierarchical universe. *Mon. Not. R. Astron. Soc.* **261**, 921–928 (1993).
- Kauffmann, J. *et al.* The COMPLETE structure of L1448: Where (and why) dense cores do form. *Astrophys. J.* (submitted).
- Pineda, J. E., Caselli, P. & Goodman, A. A. CO isotopologues in the Perseus molecular cloud complex: the X-factor and regional variations. *Astrophys. J.* **679**, 481–496 (2008).
- Bertoldi, F. & McKee, C. F. Pressure-confined clumps in magnetized molecular clouds. *Astrophys. J.* **395**, 140–157 (1992).
- Ostriker, E. C., Stone, J. M. & Gammie, C. F. Density, velocity, and magnetic field structure in turbulent molecular clouds. *Astrophys. J.* **546**, 980–1005 (2001).
- Goodman, A., Pineda, J. E. & Schnee, S. The “true” column density distribution in star-forming molecular clouds. *Astrophys. J.* (in the press); preprint at (<http://arxiv.org/abs/0806.3441v3>) (2008).
- Rosolowsky, E. W. *et al.* An ammonia spectral atlas of dense cores in Perseus. *Astrophys. J.* **175** (Suppl.), 509–521 (2008).
- Jørgensen, J. K. *et al.* Current star formation in the Ophiuchus and Perseus molecular clouds: constraints and comparisons from unbiased submillimeter and mid-infrared surveys. II. *Astrophys. J.* **683**, 822–843 (2008).
- Foster, J. B. & Goodman, A. A. Cloudshine: New light on dark clouds. *Astrophys. J.* **636**, L105–L108 (2006).

Supplementary Information is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank A. Munshi for putting us in touch with M. Thomas and colleagues at Right Hemisphere, whose software and assistance enabled the interactive PDF in this paper; P. Padoan for providing the simulated data cube; R. Shetty for comments on the paper; F. Shu for suggesting we extend our analysis to measure boundedness of  $p$ - $p$ - $v$  ‘bound’ objects in  $p$ - $p$ - $p$  space using simulations; and S. Hyman, Provost of Harvard University, for supporting the start-up of the Initiative in Innovative Computing at Harvard, which substantially enabled the creation of this work. 3D Slicer is developed by the National Alliance for Medical Image Computing and funded by the National Institutes of Health grant U54-EB005149. The COMPLETE group is supported in part by the National Science Foundation. E.W.R. is supported by the NSF AST-0502605.

**Author Contributions** The dendrogram algorithm and software was created by E.W.R. The interactive figures were assembled by M.A.B., J.K. and M.H. using software from Right Hemisphere and Adobe. J.K. and M.H. worked to allow 3D Slicer to plot the surfaces relevant to the dendrograms shown in the 3D figures. J.B.F. produced Fig. 1, and J.E.P. carried out the ‘CLUMPFINDing’ analysis shown in Fig. 2 and Supplementary Fig. 1. A.A.G. wrote most of the text, and all authors contributed their thoughts to the discussions and analysis that led to this work.

**Author Information** The 3D Slicer software used to create the surface renderings is available at <http://am.iic.harvard.edu/>. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to A.A.G. ([agoodman@cfa.harvard.edu](mailto:agoodman@cfa.harvard.edu)).



# Squeezing and over-squeezing of triphotons

L. K. Shalm<sup>1</sup>, R. B. A. Adamson<sup>1</sup> & A. M. Steinberg<sup>1</sup>

Quantum mechanics places a fundamental limit on the accuracy of measurements. In most circumstances, the measurement uncertainty is distributed equally between pairs of complementary properties; this leads to the ‘standard quantum limit’ for measurement resolution. Using a technique known as ‘squeezing’, it is possible to reduce the uncertainty of one desired property below the standard quantum limit at the expense of increasing that of the complementary one. Squeezing is already being used to enhance the sensitivity of gravity-wave detectors<sup>1</sup> and may play a critical role in other high precision applications, such as atomic clocks<sup>2</sup> and optical communications<sup>3</sup>. Spin squeezing (the squeezing of angular momentum variables) is a powerful tool, particularly in the context of quantum light–matter interfaces<sup>4–9</sup>. Although impressive gains in squeezing have been made, optical spin-squeezed systems are still many orders of magnitude away from the maximum possible squeezing, known as the Heisenberg uncertainty limit. Here we demonstrate how an optical system can be squeezed essentially all the way to this fundamental bound. We construct spin-squeezed states by overlapping three indistinguishable photons in an optical fibre and manipulating their polarization (spin), resulting in the formation of a squeezed composite particle known as a ‘triphoton’. The symmetry properties of polarization imply that the measured triphoton states can be most naturally represented by quasi-probability distributions on the surface of a sphere. In this work we show that the spherical topology of polarization imposes a limit on how much squeezing can occur, leading to the quasi-probability distributions wrapping around the sphere—a phenomenon we term ‘over-squeezing’. Our observations of spin-squeezing in the few-photon regime could lead to new quantum resources for enhanced measurement, lithography and information processing that can be precisely engineered photon-by-photon.

The standard way to characterize and represent the polarization of classical light is through the use of a Stokes vector. The vector’s components are the Stokes parameters  $S_1$ ,  $S_2$  and  $S_3$  that describe the degree of linear, diagonal and circular polarization respectively, while the intensity of the beam is represented by the Stokes parameter  $S_0$ .  $S_1$ ,  $S_2$  and  $S_3$  form a Cartesian coordinate system, and for polarized light  $S_1^2 + S_2^2 + S_3^2 = S_0^2$ . Such a Stokes vector terminates at a point on the surface of what is known as the Poincaré sphere. For quantum polarization states, a Poincaré sphere can be constructed where the Stokes parameters are replaced by the Stokes operators:

$$\begin{aligned}\hat{S}_0 &= a_H^\dagger a_H + a_V^\dagger a_V = \hat{n}_H + \hat{n}_V = N, \\ \hat{S}_1 &= a_H^\dagger a_H - a_V^\dagger a_V = \hat{n}_H - \hat{n}_V, \\ \hat{S}_2 &= a_H^\dagger a_V + a_V^\dagger a_H = \hat{n}_D - \hat{n}_A, \\ \hat{S}_3 &= i(a_V^\dagger a_H - a_H^\dagger a_V) = \hat{n}_R - \hat{n}_L,\end{aligned}\quad (1)$$

where  $N$  is the total photon number and  $a_i^\dagger$ ,  $a_i$  and  $\hat{n}_i$  are the creation, annihilation and photon number operators for polarization mode  $i$ ,

and where  $i$  takes the values H, V, D, A, R and L corresponding to horizontal, vertical, diagonal, anti-diagonal, right-hand circular and left-hand circular polarizations, respectively<sup>5–7</sup>. The Stokes operators do not commute with one another, leading to the Heisenberg uncertainty relations  $V_l V_m \geq |\langle \hat{S}_n \rangle|^2 \varepsilon_{l,m,n}$ , where  $V_l$  represents the variance in the Stokes operator  $\hat{S}_l$ . This uncertainty implies that rather than assuming definite values, the Stokes parameters must be described by a joint quasi-probability distribution. A Wigner distribution on the Poincaré sphere can be introduced following the approach of ref. 10 to provide an informationally complete description of the polarization state space, and to calculate any of the desired properties of the state<sup>11–14</sup>. It is important to note that for our triphoton states  $N$  is fixed at three and hence  $\hat{S}_0$  is a constant without fluctuations.

The state of a single polarized photon can be completely described by two parameters (relative amplitude and phase of the two orthogonal polarization modes). Because there are only two parameters, once the centre of the distribution (a point on the surface of the Poincaré sphere) is specified, no more freedom remains. All pure states are described by the same circularly symmetric quasi-probability distribution centred at a point. This is akin to the Bloch sphere for a spin-1/2 particle, where a pure state is completely identified by a Bloch vector. In general,  $N$  photons polarized in the same direction form a spin-coherent state, which is the analogue of the Glauber coherent state of a harmonic oscillator<sup>15</sup>. The quasi-probability distribution of the spin-coherent state has circular symmetry and can be represented by a single Stokes vector that terminates at the centre of the distribution, as in the single-particle case. The uncertainty in the Stokes operators of this two-level system scales as  $\sqrt{N}$ , set by binomial statistics due to the independence of measurements on individually polarized (unentangled) photons.

For a general quantum state formed by a collection of  $N$  photons in the same mode,  $2N$  degrees of freedom are required to characterize the state, and a simple vector is no longer sufficient to describe the polarization on the Poincaré sphere. In such a state, the photons do not have individual identities and it becomes impossible to speak about the polarization of a single photon. Instead, the entire system must be treated as a composite particle of spin  $N/2$  and more complex distributions that break the circular symmetry of the simple two-level system are allowed. In such distributions, polarization squeezing becomes evident as the uncertainty in one Stokes operator is decreased below the shot-noise limit of  $\sqrt{N}$  at the expense of an increase in uncertainty of the conjugate operator. Previous work on optical spin-squeezing has used such distributions, but always in a limit where the sphere could be approximated by a tangent plane<sup>5–7</sup>. In these experiments, the uncertainty distributions are squeezed by  $\sim 100 \mu\text{rad}$  on the surface of the sphere. In contrast, the squeezing we measure in our experiment is profoundly affected by the topology of the Poincaré sphere. When the degree of squeezing becomes large enough, over-squeezing occurs. When a triphoton is over-squeezed, the topology of the sphere causes the quasi-probability distribution to wrap around

<sup>1</sup>Centre for Quantum Information and Quantum Control, Institute for Optical Sciences, Department of Physics, University of Toronto, 60 St George Street, Toronto, Ontario, Canada M5S 1A7.

itself, and further squeezing leads to a counter-intuitive increase in the uncertainty of the Stokes operator being squeezed.

The triphoton states we study are of the form  $|\zeta\rangle = c_3|3,0\rangle_{H,V} + c_2|2,1\rangle_{H,V} + c_1|1,2\rangle_{H,V} + c_0|0,3\rangle_{H,V}$ , where  $|m,n\rangle_{H,V}$  represents  $m$  and  $n$  photons in the orthogonal horizontal (H) and vertical (V) polarization modes respectively. To experimentally produce these triphoton states, we place a pair of photons generated by type-II spontaneous parametric downconversion (SPDC) and a single photon from an attenuated local oscillator (LO) into the same spatial, temporal and spectral mode following the method of refs 16–18, as shown in Fig. 1. Because the photons do not come from true single-photon sources, unavoidable background terms occur, as discussed further in the Supplementary Information. To minimize background events, the pair generation rate and the local oscillator rate are carefully balanced to ensure that the majority of three-photon events detected originate from a single SPDC pair and a single LO photon. Full quantum state tomography is carried out on the triphoton states using the hidden differences technique<sup>19,20</sup>, and the resulting density matrix used to calculate the Wigner quasi-probability distributions on the Poincaré sphere (as discussed further in the Supplementary Information). A total of 11 different triphoton states were prepared and measured as shown in Fig. 1 by varying the squeezing factor  $T$  from 0 to 1.7 (see Methods for further details). Here  $T = T_V/T_H$ , the ratio of the transmissivity of the horizontal ( $T_H$ ) and the vertical ( $T_V$ ) modes of the variable partial polarizer. The Wigner quasi-probability distributions for five of these states are shown in Fig. 2a.

When  $T = 0$ , the partial polarizer only transmits the horizontal polarization, producing a quasi-classical coherent state composed of three horizontally polarized photons. A quarter-wave plate then rotates the coherent state into the circular basis. On the Poincaré sphere, this coherent state has a circularly symmetric quasi-probability distribution centred on the  $\hat{S}_3$  axis (as shown in Fig. 2b), and an uncertainty in  $\hat{S}_1$  and  $\hat{S}_2$  given by the shot-noise limit.

Setting  $T = 0.7$  produces the phase state  $|\zeta\rangle = \frac{1}{\sqrt{2}}(|3,0\rangle_{H,V} + |2,1\rangle_{H,V} + |1,2\rangle_{H,V} + |0,3\rangle_{H,V})$ , which represents an equally weighted superposition of polarization states. On the Poincaré sphere, the Wigner distribution is centred on the  $\hat{S}_3$  axis (as shown in Fig. 2c), but now takes a ‘banana’ shape as the uncertainty along  $\hat{S}_2$  decreases below the shot-noise limit while the uncertainty in  $\hat{S}_1$  increases above it. The phase state has been called a ‘jack-of-all-trades’ state<sup>21</sup> for metrology, as its symmetry and squeezing properties allow it to accurately acquire and then track an unknown phase.

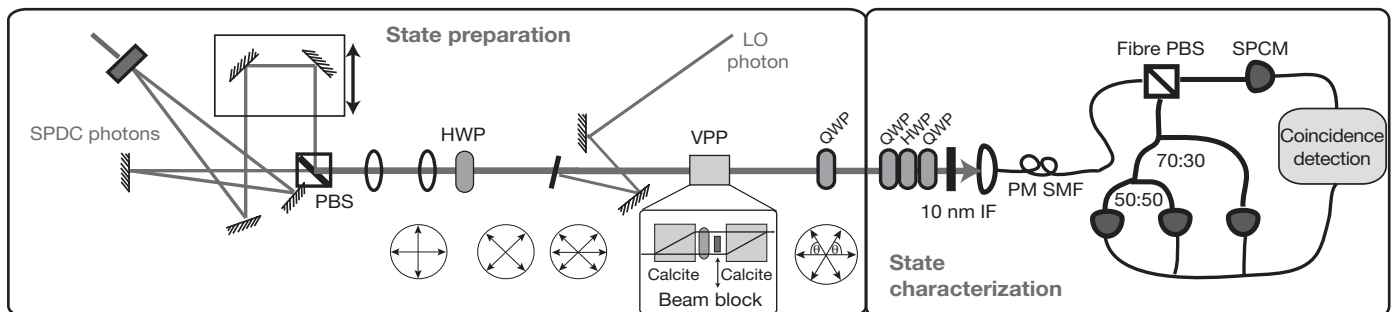
Finally, tuning the partial polarizer to  $T = 1.7 \approx \sqrt{3}$  creates the state  $|\zeta\rangle = \frac{1}{\sqrt{2}}(|N,0\rangle_{H,V} + |0,N\rangle_{H,V})$ , where for us  $N=3$ . Such states are called ‘N00N’ states and are capable of reaching the Heisenberg limit in phase-sensitive measurements and also of achieving super-resolution in lithographic applications<sup>16,22–25</sup>. Our N00N state has a

fidelity of  $0.68 \pm 0.01$  with an ideal N00N state. This low fidelity stems from background events in our state preparation and the fact that the three photons are not completely indistinguishable. The major background contributions are due to either two SPDC pairs or three LO photons triggering a threefold coincidence. To account for the background contributions, state tomography was independently carried out on each major background source (by alternately blocking the LO and SPDC photons) and then subtracted off the measured density matrix, yielding a fidelity with an ideal N00N state of  $0.80 \pm 0.01$ . This background-subtracted N00N state still contains smaller background contributions that make up  $\sim 10\%$  of the measured state (as further discussed in the Supplementary Information). Because the three photons are not generated by the same process, there is also unavoidable distinguishing information present in the state. This distinguishing information requires more parameters, and therefore a larger density matrix, to fully specify the polarization of the state. The extra elements in this density matrix represent the entanglement of the triphoton state with the environment, and provide a lower bound on the amount of distinguishing information present<sup>19</sup>. With the development of ideal single-photon sources, both the unwanted background contributions and the residual distinguishing information in our states will be eliminated and our protocol can be scaled beyond three photons.

In the metrology and lithography applications for which N00N states have been proposed, the detection schemes rely on  $N$ -photon absorbers that are sensitive only to symmetric states. Because the extra elements in the density matrix resulting from distinguishing information represent states of mixed symmetry, they will be automatically filtered out by the detectors in any practical implementation. After performing such filtering and accounting for the major background contributions (as further discussed in the Supplementary Information), our N00N state has a fidelity of  $0.89 \pm 0.01$  with an ideal state. As shown in Fig. 2d, the Wigner quasi-probability distribution is peaked at the poles of the  $\hat{S}_1$  axis, corresponding to finding all three photons being either horizontally or vertically polarized, and displays a threefold symmetry around the equator. This threefold symmetry is directly related to the threefold enhancement in phase super-resolution of this N00N state<sup>16</sup>, and the minimum feature width corresponds to the Heisenberg-limited phase super-sensitivity possible in interferometry using N00N states<sup>24,26</sup>.

To quantify the degree of entanglement in our states, we calculate the negativity  $N(\rho) = \sum \max(0, \mu_j)$ , a measure related to the Peres–

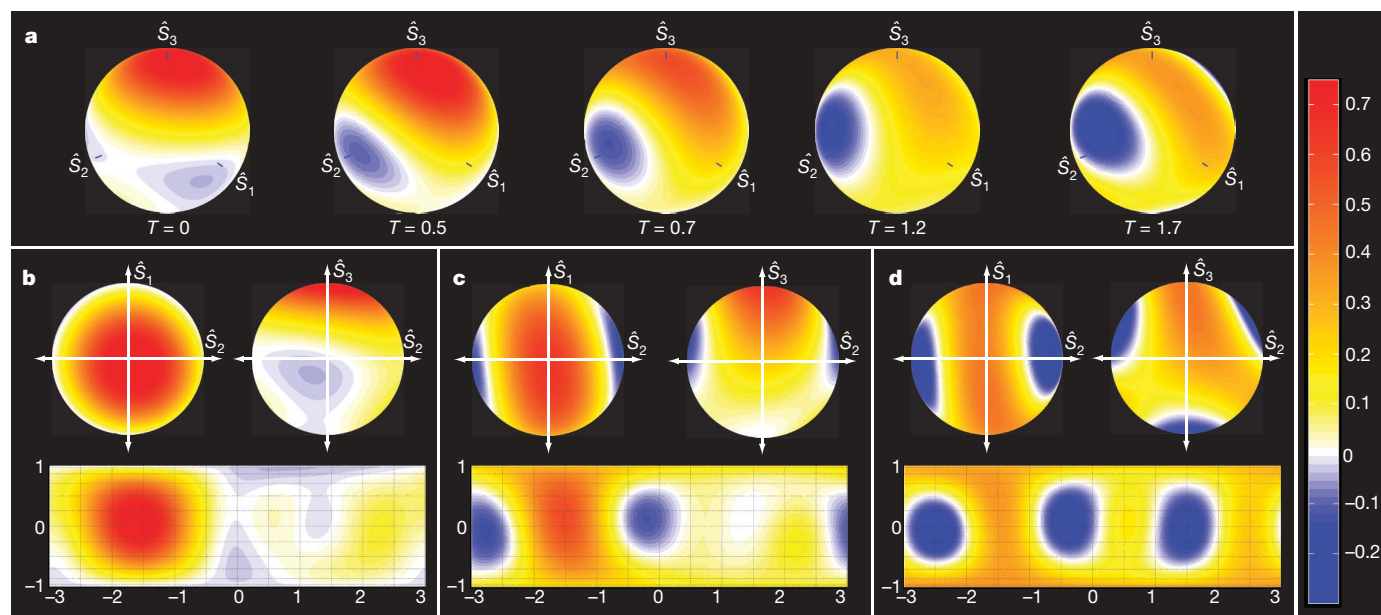
Horodecki criterion<sup>9,27–29</sup>, where the  $\mu_j$ s are the eigenvalues of the partial transpose  $\rho^{\text{PT}}$  of the density matrix  $\rho$ . Any positive value of the negativity indicates entanglement, with maximally entangled states such as the N00N state having  $N(\rho) = 1$ . Using the negativity, we can determine the bipartite entanglement between any single



**Figure 1 | Overview of the experiment.** A pair of downconverted photons from type-II spontaneous parametric downconversion (SPDC) are placed into the same mode with a polarizing beamsplitter (PBS). A half-wave plate (HWP) rotates the photons into the diagonal basis, and a third photon from a local oscillator (LO) is combined into the same mode using a laser window. A variable partial polarizer (VPP) is used to manipulate the polarization of

the photons. The resulting triphotons are coupled into a single mode fibre (SMF) that is polarization maintaining (PM) and spectrally filtered using an interference filter (IF). Quantum state tomography is carried out by using a polarization analyser consisting of two quarter-wave plates (QWP), one HWP and a PBS, and threefold coincidence counts are collected using coincidence circuitry and single-photon counting modules (SPCMs).





**Figure 2 | Triphoton Wigner quasi-probability distributions on the Poincaré sphere.** **a**, Wigner quasi-probability distributions for five triphoton states with different squeezing parameters  $T$ . **b–d**, Two-dimensional projections (top) and an equal-area cylindrical map projection (bottom) of the quasi-probability distributions for a coherent state (**b**), phase state (**c**) and a background-subtracted N00N state (**d**). In **b**, the coherent

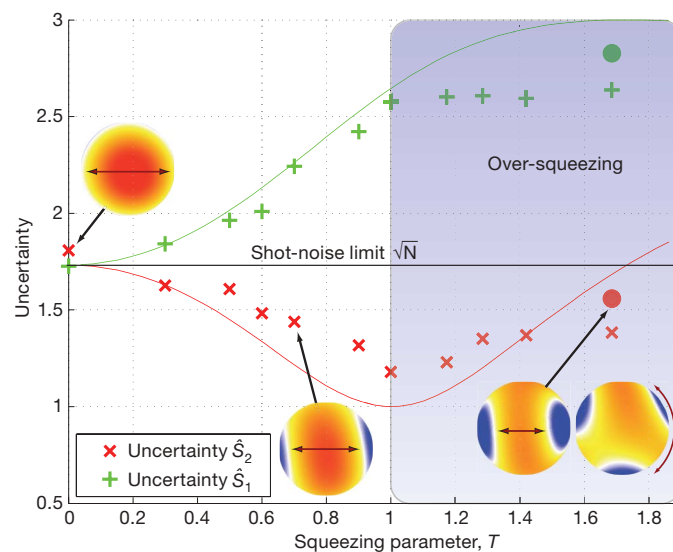
state has a uniform distribution centred about the  $\hat{S}_3$  axis. As the squeezing parameter is increased to produce a phase state, the width of the angular features decreases until the minimum feature size is reached in **d** with the maximally-squeezed N00N state. In both the phase state and N00N state, large regions of quasi-probability distribution take on negative values. The colour scale at right applies to all panels.

photon and the rest of the system. Starting with the coherent state at  $T = 0$  with a negativity of  $0.05 \pm 0.03$ , the degree of entanglement steadily increases until the background-corrected N00N state is reached, which has a negativity of  $0.52 \pm 0.03$ .

The uncertainty in the Stokes operators is measured for each of the 11 states created, as shown in Fig. 3. Starting with the coherent state at  $T = 0$  and continuing to the state with  $T = 1$ , the uncertainty in  $\hat{S}_2$  decreases below the shot-noise limit while the degree of squeezing increases. However, after  $T = 1$ , the states begin to over-squeeze, as further squeezing leads to an increase in the uncertainty of both  $\hat{S}_1$  and  $\hat{S}_2$ . As the quasi-probability distributions wrap around the surface of the Poincaré sphere, the three photons become more entangled but appear to depolarize with respect to the Stokes operators. The N00N state is maximally entangled and contains perfect correlations, yet appears completely unpolarized to intensity measurements. Unlike mixed states, which appear as broad probability distributions on the surface of the Poincaré sphere (with the maximally mixed state represented by a uniform distribution), the N00N state distribution contains the sharpest features possible due to the quantum correlations that exist. This is analogous to the maximally entangled Greenberger–Horne–Zeilinger (GHZ) state where any single member appears mixed, yet perfect correlations exist between all particles. For both the N00N and the GHZ states, these correlations only become evident when performing multi-particle measurements using, for example, number-resolving detectors.

In our experiment, the presence of background terms leads to a reduction in the purity and amount of squeezing possible. As discussed in the Supplementary Information, the background terms are polarized and add incoherently to the desired triphoton state. It is interesting to note that the background-subtracted N00N state has a larger uncertainty in  $\hat{S}_2$  than the non-subtracted case. This is because the background contributions in the state at  $T = 1.7$  reduce the amount of squeezing present, but the state is still within the over-squeezing regime and therefore a decrease in uncertainty is measured. Once the dominant background contributions are experimentally measured and removed, the purity increases and the quantum correlations become stronger, leading to an increase in the uncertainty of

$\hat{S}_2$ . When higher order moments are measured, the background-subtracted N00N state exhibits phase resolution and sensitivity superior to those of the non-background-subtracted state. Squeezed states



**Figure 3 | The uncertainty in the Stokes parameters  $\hat{S}_1$  and  $\hat{S}_2$  for 11 squeezed triphoton states.** As the squeezing parameter is increased to  $T = 1$ , the uncertainty in  $\hat{S}_2$  decreases. After  $T = 1$ , the topology of the sphere causes the triphoton states to over-squeeze, and the uncertainty in  $\hat{S}_2$  begins to increase again as the state becomes depolarized with respect to single-particle intensity measurements. Insets show the Wigner quasi-probability distributions for the spin-coherent state  $T = 0$  (top left), the phase state  $T = 0.7$  (bottom left), and the background-corrected N00N state  $T = 1.7$  (bottom right). For the N00N state, both the side and top views are shown to illustrate how the over-squeezing leads to an increase in the uncertainty of  $\hat{S}_2$ . Solid lines indicate theory, crosses are experimental data points. The latter each have an error of less than 2%. The discrepancy between the measured states and theory is due to background contributions. Filled circles are the uncertainties in the N00N state after corrections for background events and distinguishing information have been made.

are often studied in the context of improving phase measurements with sub-shot-noise sensitivity. However, for highly squeezed states in the over-squeezing limit (such as the N00N state), multi-photon detectors as opposed to intensity-based measurements are required to take advantage of the enhanced precision offered.

This is the first time that spin-squeezing in the few-photon limit has been observed. Recently, there have also been studies of atomic spin-squeezed states which implicitly contain over-squeezing<sup>8,9</sup>, but our work represents the first time that the phenomenon of over-squeezing and its relation to quantum metrology has been directly investigated. It unifies the concepts of spin-squeezing of intense beams and multi-photon Heisenberg-limited metrology, highlighting the deep connection between squeezing and entanglement. This is indicated in the maximally squeezed N00N state, where no single-particle information remains and all the information resides purely in non-classical correlations. With improvements in single-photon sources, this relationship between squeezing and entanglement could be further exploited to engineer custom non-classical states that would advance key quantum technologies, including quantum computation, lithography and cryptography.

## METHODS SUMMARY

To produce our triphoton states, we start with a pulsed 40 MHz Ti:sapphire laser operating at 810 nm, which produces 100-fs pulses that are then frequency-doubled using a 0.5-mm-thick  $\beta$ -barium borate (BBO) crystal to produce 50 mW of pulsed light at 405 nm. These pulses are focused into a 0.5-mm BBO crystal, phase-matched for type-II spontaneous parametric downconversion with beam-like emission<sup>30</sup> to generate a pair of photons that are combined into the same mode using a polarizing beamsplitter. These two photons are then sent through a half wave-plate at 22.5° to produce the state  $a_{45}^\dagger a_{-45}^\dagger |0\rangle = (a_H^{\dagger 2} - a_V^{\dagger 2})|0\rangle$  where H and V are the horizontal and vertical polarization modes, respectively. Next, the downconverted pair of photons is combined with a horizontally polarized laser photon on a highly transmissive beamsplitter to produce the unnormalized state  $a_{45}^\dagger a_{-45}^\dagger a_H^\dagger |0\rangle = (a_H^{\dagger 2} - a_V^{\dagger 2})a_H^\dagger |0\rangle$ . This state is then sent to a variable partial polarizer where either the horizontal or vertical polarization component can be attenuated and the polarization state of the three photons controlled. The action of the variable partial polarizer is to squeeze the polarization distribution of the photons, producing the state  $|\zeta\rangle = \eta e^{-\delta_1 \ln(T)} a_{45}^\dagger a_{-45}^\dagger a_H^\dagger |0\rangle$ , where  $\eta$  is a normalization factor and  $T = T_V/T_H$  is the squeezing factor given by the ratio of the transmissivity of the horizontal  $T_H$  and the vertical  $T_V$  modes of the partial polarizer. Following the partial polarizer, a quarter-wave plate is used to rotate the state from the circular to linear polarization basis. A 12 nm spectral filter centred at 810 nm and a single-mode fibre are used to remove spectral and spatial distinguishing information.

A polarization analyser, consisting of a quarter-wave plate followed by a fibre polarization beamsplitter and four single-photon detectors, is used to carry out quantum state tomography as further discussed in the Supplementary Information. It is impossible to combine photons with non-orthogonal polarizations into the same mode with certainty; therefore we use post-selection to study only the cases where all three photons are detected in the same mode.

Received 13 August; accepted 4 November 2008.

1. Goda, K. *et al.* A quantum-enhanced prototype gravitational-wave detector. *Nature Phys.* **4**, 472–476 (2008).
2. Ye, J., Kimble, H. & Katori, H. Quantum state engineering and precision metrology using state-insensitive light traps. *Science* **320**, 1734–1738 (2008).
3. Furusawa, A. *et al.* Unconditional quantum teleportation. *Science* **282**, 706–709 (1998).

4. Hald, J., Sørensen, J. & Polzik, E. Spin squeezed atoms: A macroscopic entangled ensemble created by light. *Phys. Rev. Lett.* **83**, 1319–1322 (1999).
5. Bowen, W., Schnabel, R., Bachor, H.-A. & Lam, P. Polarization squeezing of continuous variable Stokes parameters. *Phys. Rev. Lett.* **88**, 093601 (2002).
6. Heersink, J. *et al.* Polarization squeezing of intense pulses with a fiber-optic Sagnac interferometer. *Phys. Rev. A* **68**, 013815 (2003).
7. Marquardt, C. *et al.* Quantum reconstruction of an intense polarization squeezed optical state. *Phys. Rev. Lett.* **99**, 220401 (2007).
8. Chaudhury, G. *et al.* Quantum control of the hyperfine spin of a Cs atom ensemble. *Phys. Rev. Lett.* **99**, 163002 (2007).
9. Ghose, S. S., Stock, R., Jessen, P. S., Lal, R. & Silberfarb, A. Chaos, entanglement and decoherence in the quantum kicked top. *Phys. Rev. A* **78**, 042318 (2008).
10. Agarwal, G. Relation between atomic coherent-state representation, state multipoles, and generalized phase-space distributions. *Phys. Rev. A* **24**, 2889–2896 (1981).
11. Dowling, J., Agarwal, G. & Schleich, W. Wigner distribution of a general angular-momentum state: Applications to a collection of two-level atoms. *Phys. Rev. A* **49**, 4101–4109 (1994).
12. Stratonovich, R. L. Gauge invariant generalization of Wigner distributions. *Dokl. Akad. Nauk SSSR* **109**, 72–75 (1956).
13. Luis, A. Quantum polarization distributions via marginals of quadrature distributions. *Phys. Rev. A* **71**, 053801 (2005).
14. Luis, A. Nonclassical polarization states. *Phys. Rev. A* **73**, 063806 (2006).
15. Arecchi, F., Courtens, E., Gilmore, R. & Thomas, H. Atomic coherent states in quantum optics. *Phys. Rev. A* **6**, 2211–2273 (1972).
16. Mitchell, M. W., Lundeen, J. S. & Steinberg, A. M. Super-resolving phase measurements with a multiphoton entangled state. *Nature* **429**, 161–164 (2004).
17. Lee, H., Kok, P., Cerf, N. J. & Dowling, J. P. Linear optics and projective measurements alone suffice to create large photon-number path entanglement. *Phys. Rev. A* **65**, 030101 (2002).
18. Fiurásek, J. Conditional generation of n-photon entangled states of light. *Phys. Rev. A* **65**, 053818 (2002).
19. Adamson, R. B. A., Shalm, L. K., Mitchell, M. W. & Steinberg, A. M. Multiparticle state tomography: hidden differences. *Phys. Rev. Lett.* **98**, 043601 (2007).
20. Adamson, R. B. A., Turner, P. S., Mitchell, M. W. & Steinberg, A. M. Detecting hidden differences via permutation symmetries. *Phys. Rev. A* **78**, 033832 (2008).
21. Durkin, G. A. & Dowling, J. P. Local and global distinguishability in quantum interferometry. *Phys. Rev. Lett.* **99**, 070801 (2007).
22. Dowling, J. Correlated input-port, matter-wave interferometer: Quantum-noise limits to the atom-laser gyroscope. *Phys. Rev. A* **57**, 4736–4746 (1998).
23. Walther, P. *et al.* De Broglie wavelength of a non-local four-photon state. *Nature* **429**, 158–161 (2004).
24. Nagata, T., Okamoto, R., O'Brien, J., Sasaki, K. & Takeuchi, S. Beating the standard quantum limit with four-entangled photons. *Science* **316**, 726–729 (2007).
25. Ou, Z. Fundamental quantum limit in precision phase measurement. *Phys. Rev. A* **55**, 2598–2609 (1997).
26. Resch, K. J. *et al.* Time-reversal and super-resolving phase measurements. *Phys. Rev. Lett.* **98**, 223601 (2007).
27. Miranowicz, A. & Grudka, A. A comparative study of relative entropy of entanglement, concurrence and negativity. *J. Opt. B* **6**, 542–548 (2004).
28. Peres, A. Separability criterion for density matrices. *Phys. Rev. Lett.* **77**, 1413–1415 (1996).
29. Horodecki, M. Separability of mixed states: necessary and sufficient conditions. *Phys. Lett. A* **223**, 1–8 (1996).
30. Takeuchi, S. Beamlike twin-photon generation by use of type II parametric downconversion. *Opt. Lett.* **26**, 843–845 (2001).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank S. Ghose and P. Jessen for discussions. This work was supported by the Natural Sciences and Engineering Research Council of Canada, Ontario Centres of Excellence, Canadian Institute for Photonic Innovations, Quantum Works, and the Canadian Institute for Advanced Research.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to L.K.S. (lshalm@physics.utoronto.ca).



# Optical manipulation of nanoparticles and biomolecules in sub-wavelength slot waveguides

Allen H. J. Yang<sup>1</sup>, Sean D. Moore<sup>2</sup>, Bradley S. Schmidt<sup>3</sup>, Matthew Klug<sup>2</sup>, Michal Lipson<sup>3</sup> & David Erickson<sup>2</sup>

The ability to manipulate nanoscopic matter precisely is critical for the development of active nanosystems. Optical tweezers<sup>1–4</sup> are excellent tools for transporting particles ranging in size from several micrometres to a few hundred nanometres. Manipulation of dielectric objects with much smaller diameters, however, requires stronger optical confinement and higher intensities than can be provided by these diffraction-limited<sup>5</sup> systems. Here we present an approach to optofluidic transport that overcomes these limitations, using sub-wavelength liquid-core slot waveguides<sup>6</sup>. The technique simultaneously makes use of near-field optical forces to confine matter inside the waveguide and scattering/adsorption forces to transport it. The ability of the slot waveguide to condense the accessible electromagnetic energy to scales as small as 60 nm allows us also to overcome the fundamental diffraction problem. We apply the approach here to the trapping and transport of 75-nm dielectric nanoparticles and  $\lambda$ -DNA molecules. Because trapping occurs along a line, rather than at a point as with traditional point traps<sup>7,8</sup>, the method provides the ability to handle extended biomolecules directly. We also carry out a detailed numerical analysis that relates the near-field optical forces to release kinetics. We believe that the architecture demonstrated here will help to bridge the gap between optical manipulation and nanofluidics.

In a series of recent articles, Psaltis *et al.*<sup>9</sup> and Monat *et al.*<sup>10</sup> describe the enormous potential of integrating optical and microfluidic elements into 'lab-on-a-chip' devices, particularly in improving fluid and particle manipulations. Traditionally accomplished through direct particle manipulation with laser tweezers<sup>1–3,11</sup> or indirectly using optically induced microfluidic effects<sup>12,13</sup>, the precision with which particles can be manipulated with these techniques makes them particularly useful for applications ranging from flow cytometry<sup>3,14</sup> to self-assembly<sup>15</sup>. Fundamentally, however, these free-space systems are limited in two ways. First, diffraction limits how tightly the light can be focused and thereby limits the overall strength of the trap. Second, the trapping region has a very short focal depth, preventing the continuous transport of nanoparticles by means of radiation pressure.

To improve trapping stability, a number of near-field methods have recently been developed<sup>16–18</sup>. Examples are the use of interfering Gaussian beams reflected off a prism surface to sort 350-nm polystyrene beads<sup>19</sup>, and the use of localized plasmonic resonances in surface-bound metallic nanostructures to trap 200-nm dielectric particles<sup>4</sup>. Waveguide-based optical transport<sup>20,21</sup> is analogous to these near-field methods in that the evanescent field extending into the surrounding liquid serves to attract particles to the waveguide. However, particles also experience photon scattering and absorption forces which propel them along it for a distance limited only by the losses in the system. Recent efforts in this area have demonstrated the sustained propulsion of dielectric microparticles<sup>20,22</sup>, metallic nanoparticles<sup>23,24</sup> and cells<sup>25</sup>. The limitation that prevents these systems

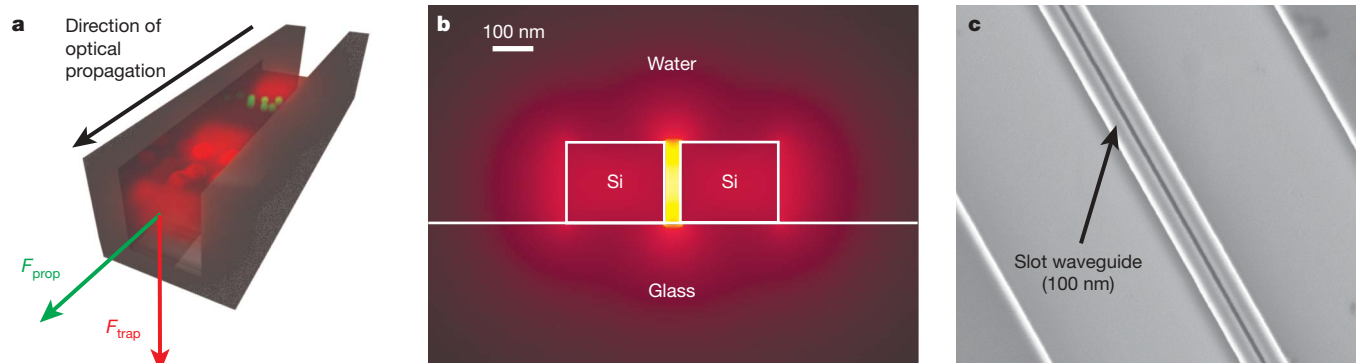
from manipulating smaller matter, including biomolecules, is that the particles only interact with the small portion of total transported light because most of it is confined within the solid core of the waveguide.

Recently Almeida *et al.*<sup>6</sup> developed a nanophotonic structure, known as a slot waveguide, that can overcome this challenge. Illustrated in Fig. 1, the slot waveguide comprises a nanoscale slot sandwiched between two materials of much higher refractive index. As shown in Fig. 1b, there exists a pseudo-transverse-electric (TE) mode that has a large electric field discontinuity at the horizontal boundaries of the slot region. The result is a high-intensity eigenmode in the slot making the majority of the optical energy accessible within the low-index fluid region. Here we demonstrate the use of these sub-wavelength-scale slot waveguides for optical capture, trapping and transport of both dielectric nanoparticles and  $\lambda$ -DNA molecules. We can achieve stable trapping of particles as small as 75 nm, representing some of the smallest dielectric matter ever trapped or transported using such a system. In addition to the experimental results, we also present an examination of the effect that the presence of a particle has on the optical mode, details of the trap strength and stiffness in comparison with other techniques, and a unique stability analysis descriptive of the release kinetics for particles near the stability point.

As shown in Fig. 2a (see also Supplementary Movies 1 and 2), we are able to capture and stably trap polystyrene nanoparticles (refractive index  $n = 1.45$ ) with diameters of 75 nm and 100 nm in slot waveguides with widths of 100 nm and 120 nm, respectively. In all cases the optical power at the exit of the fibre used to couple light into the waveguide was less than 300 mW, the excitation wavelength was  $\lambda = 1,550$  nm, and the trapping used TE polarization. Experimental details are available in the Supplementary Information. The first movie (Supplementary Movie 1) illustrates our ability to capture and accumulate flowing particles in the slot waveguide for indefinite periods of time and release them by either reducing the optical power or switching the polarization. Excitation of the slot waveguides using transverse-magnetic (TM) polarization required three to five times as much power to obtain stable trapping, so switching polarization also tended to break the trap. The microfluidic flow serves to transport the particles to the waveguide but does not have a role in the trapping itself. This is indicated by the fact that the trap breaks on removal of the optical excitation in the above experiments.

The second movie (Supplementary Movie 2) illustrates the dynamics of the capture of flowing particles in a trap near the stability point, where the random thermal energy in the system is of the same order as the amount of work required to break the trap<sup>26</sup>. The average retention time in such a trap is a statistical process governed by the release kinetics of the system, which in turn is governed by the trap strength, its stiffness and the location on the waveguide where the particle is trapped (as will be described below). In the experiment

<sup>1</sup>School of Chemical and Biomolecular Engineering, <sup>2</sup>Sibley School of Mechanical and Aerospace Engineering, <sup>3</sup>School of Electrical and Computer Engineering, Cornell University, Ithaca, New York 14853, USA.



**Figure 1 | Nanoscale optofluidic transport.** **a**, Schematic illustrating the transport of two different sizes of nanoparticles in a slot waveguide. The force  $F_{\text{prop}}$  represents the radiation pressure force responsible for optofluidic transport, and  $F_{\text{trap}}$  is the trapping force that holds nanoparticles within the slot region. **b**, Mode profile for a silicon-on-insulator 40-nm slot waveguide

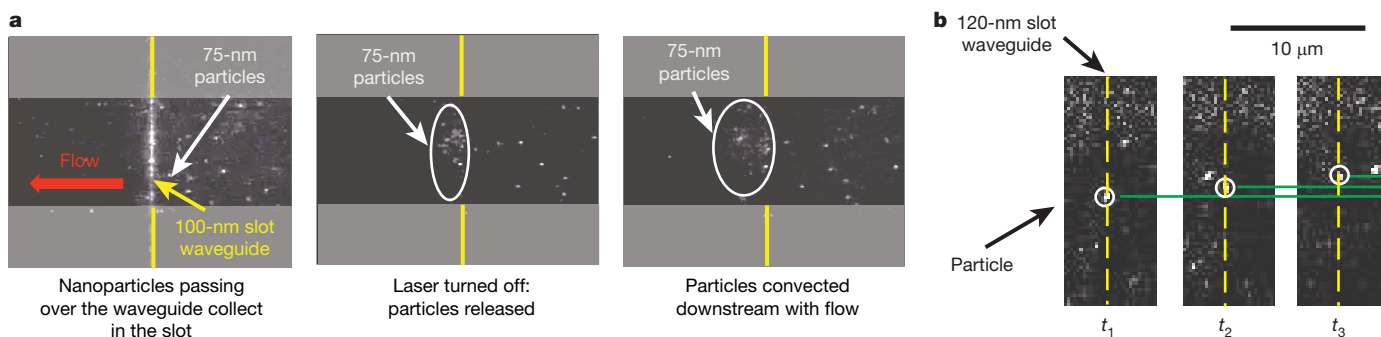
shown, the nanoparticles are flowing by in the microchannel at an average speed of  $80 \mu\text{m s}^{-1}$  and, as can be seen, the number of particles that are captured is low (less than 25%) in comparison with those that flow by. The reason for the low capture rate is that to be trapped, a particle must be on a streamline that passes through the evanescent field. This is analogous to a situation in which a flowing particle must pass through the focal point of a free-space optical tweezer in order to be trapped, and is not an inherent limitation of our system. The capture rate can be increased by decreasing the channel size, reducing the flow rate or increasing the optical power.

As mentioned above, an advantage of this approach is the ability not only to capture nanoscopic matter but to transport it optically. This capability is important for the development of active nano-assembly techniques and for optically driven bioanalytics<sup>27</sup>. From Rayleigh theory it is well known that the radiation-pressure-based transport velocity of a dielectric nanoparticle is proportional to the local intensity and scales with the fifth power of particle radius<sup>28</sup>. As such it is extremely difficult to optically transport very small matter unless very high optical intensities can be achieved. As shown in Fig. 2b (and Supplementary Movie 3), using our slot waveguide system we have been able to demonstrate optical propulsion of 100-nm particles at an average speed of  $1.5 \mu\text{m s}^{-1}$  (using 250 mW optical power measured at the exit of the fibre). Because the propulsion velocity is inversely proportional to the fourth power of wavelength, one method by which the transport velocity could be increased is by using a different high-refractive-index material that is transparent at lower wavelengths.

immersed in water, calculated using a finite-element simulation package. The main trapping region is in the high-intensity slot mode, although alternate trapping locations are located on the sides of the waveguide, where there are two decaying evanescent modes. **c**, Scanning electron microscope image of 100-nm slot waveguide structure

Nanoscale dielectric particles can be considered as coarse approximate models for biological species such as viruses and very small bacteria. Of perhaps greater interest is the ability to capture and optically confine individual biomolecules. As shown in Fig. 3 (and Supplementary Movie 4) we have been able to capture from solution and stably trap individual strands of YOYO-1 tagged  $\lambda$ -DNA molecules 48 kilobases long. As in previous cases, trapping was accomplished using 250 mW of input optical power at a wavelength of 1,550 nm; in this case, however, we used a 60-nm slot waveguide. As can be seen in the second half of the movie, when the power is removed the DNA strands are released. These experiments were done under pH conditions in which the DNA is reported to be in a partially extended state<sup>7</sup>. Although others have demonstrated optical trapping of  $\lambda$ -DNA at pH levels at which the molecule is known to be in a supercoiled state<sup>7,8</sup>, it has proved difficult to trap partially extended molecules because the focal point of a tightly focused tweezer can only interrogate a small portion of the molecule. The slot waveguide technique allows us to trap extended molecules, as the confinement force is equivalently applied along a line rather than at a point. Further development of the transport technique may also allow us to develop new biomolecular separation mechanisms or new methods of interrogating single molecules for rapid sequencing<sup>29</sup> or direct haplotyping<sup>30</sup>.

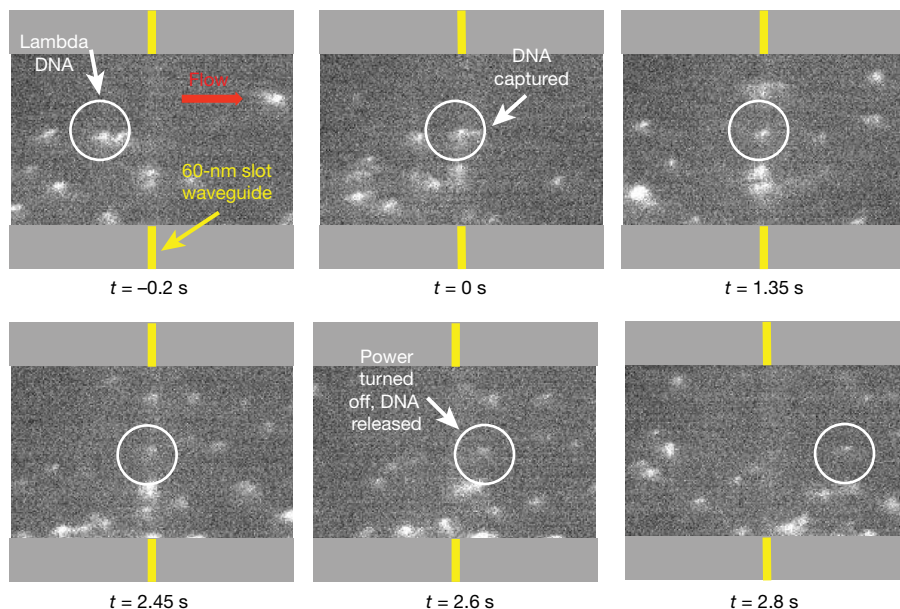
To characterize the trapping stability, stiffness and release kinetics better, we have carried out a detailed three-dimensional numerical analysis of the system. Referring back to Fig. 2, although most of the particles were observed to be trapped in the high-intensity slot region, trapping was also observed along the sides of the slot waveguide



**Figure 2 | Trapping and transport of dielectric nanoparticles in a slot waveguide.** Movies illustrating nanoparticle capture, transport and release are included in the Supplementary Information. **a**, Waveguide is optically excited while 75-nm polystyrene nanoparticles flow in the overlying microchannel over 100-nm slot waveguides. Over time, particles collect in the slot and also on the sides of the waveguide. At  $t = 0$ , the laser source is removed and particles are released from the waveguide. Immediately after

release, a 'cloud' of particles forms as the particles leave their trapping sites and the released particles are carried down the channel by the fluid flow. **b**, Trapped 100-nm nanoparticles in 120-nm slot waveguides are transported a short distance by radiation pressure. Time-lapse images are cropped from images taken using a SensiCam CCD camera with contrast and brightness adjustments to the entire image. The cropping location is the same in each time-lapse image.

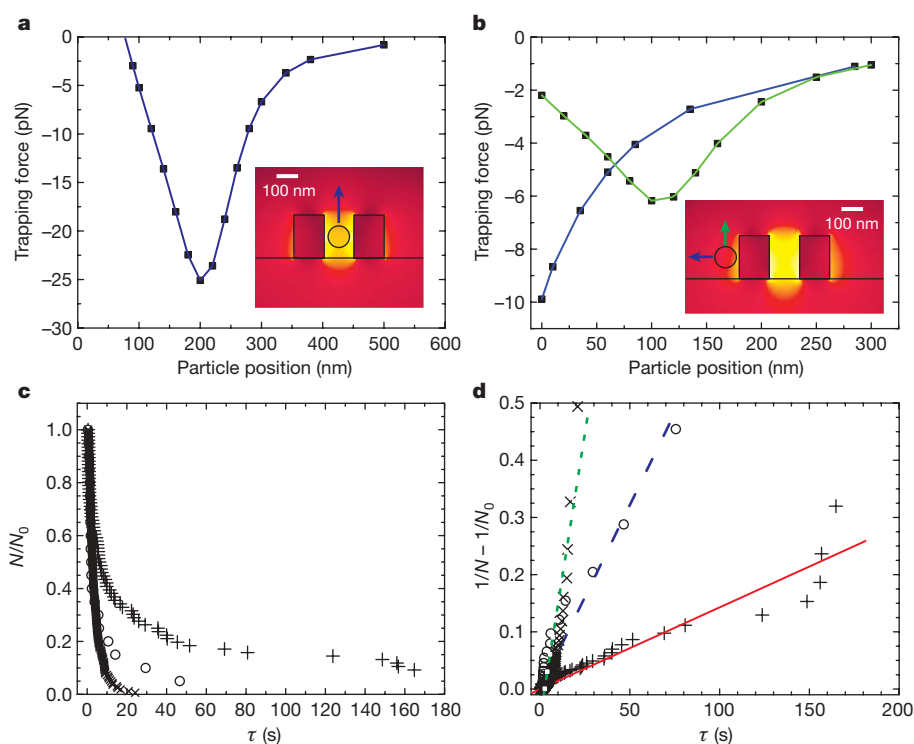


**Figure 3 | Capture and trapping of  $\lambda$ -DNA.**

Movies illustrating the DNA capture and release are included in the Supplementary Information. Images show individual YOYO-tagged 48-kilobase  $\lambda$ -DNA flowing over an optically excited 60-nm-wide slot waveguide. At time  $t = 0$  the encircled DNA is trapped. In this case the DNA is released at the 2.6-s mark and flows downstream. Supplementary Movie 4 shows the collection of DNA molecules over time and their release in response to removal of the optical excitation. Trapping conditions are identical to those used for the nanoparticle experiments shown in Fig. 2. Time-lapse images are cropped from images taken using a SensiCam CCD camera with contrast and brightness adjustments to the entire image. The cropping location is the same in each time-lapse image.

structure. Because the trap strength is related to the local optical field intensity, we can compare the behaviours of particles trapped in the two different regions to gauge the effect on trapping stability. To do this, we used a finite-element simulation to calculate the relative

trapping force for a side-trapped particle in comparison with one trapped in the slot, shown in Fig. 4a, b. As can be seen, the trapping force for side-trapped particles is much smaller, and as a result less work energy is required to exceed the trapping release barrier. This



**Figure 4 | Trapping forces on particles trapped inside and outside the slot waveguide, and analysis of particle release kinetics.** The work required to release a particle from the trap is found by integrating the force curves from a particle's stable trapping position to infinity. Inset images show the calculated electric field distribution with the particle in the two stable trapping positions. The arrows indicate the direction of release. **a**, Trapping force for a 100-nm particle in a 120-nm slot plotted against its position relative to the height of the slot waveguide. The particle position is measured as the distance of the particle centre from the bottom of the 200-nm-tall slot waveguide. The trapping force reaches a maximum at the point where the field gradient is strongest at the entrance to the slot waveguide. **b**, Trapping force for a 100-nm particle side-trapped in the same structure with two possible release paths. Particle position is measured as the deviation of the

particle centre from its most stable position. The blue line corresponds to a particle being pushed off the waveguide by an external force, and the green line corresponds to a particle being lifted off the waveguide. The trapping forces listed are normalized to 1 W of guided power. **c**, Experimental data for  $N/N_0$  plotted against normalized time,  $\tau$ . Particles trapped in the slot are represented by crosses, side-trapped particles are represented by circles, and DNA by  $\times$  signs. The quantity  $N/N_0$  represents the relative 'concentration' of particles trapped on the waveguide. **d**, Plot of  $F(N)$  (see Supplementary Information) for a second-order rate law ( $1/N - 1/N_0$ ) versus  $\tau$ . The same symbols are used here as in **c**. The lines represent linear fits to the data, from which the release rate constant can be obtained. In this case sharper slopes represent a higher release rate and therefore a less stable trap.

provides an avenue to differentiate the two situations, as we would expect side-trapped particles to be released more easily than their slot-trapped analogues. Figure 4a, b also illustrates that the optical mode of the slot waveguide is not greatly perturbed by the presence of the particle in either position.

Because the direction of trap release is in the vertical direction, and because of the physical confinement provided by the channel walls and the fact that the trapping force does not vary along the length of the waveguide, it is difficult to extract quantitative values for the trapping stability experimentally. From the numerical computations, however, we can estimate the trap stiffness from the slope of the force-distance curve as being 0.2 pN per nanometre for a 100-nm nanoparticle at 1-W excitation power. Although a direct comparison is difficult, this is significantly higher than previously described for larger particles using other near-field techniques (see Grigorenko *et al.*<sup>4</sup> who reported a 0.013 pN nm<sup>-1</sup> stiffness for a 200-nm bead).

As described in previous analytical work<sup>26</sup>, for a particle in an optical near field, there is a finite amount of work energy required to remove a particle from a stably trapped location to one where the trap no longer has any influence. When trapping is relatively weak and the particles small, the random thermal energy in the system will eventually exceed this work and the particle will be released. In such cases, the work energy is analogous to an activation energy barrier that impedes the release of particles from the waveguide, and the kinetic behaviour is similar to molecular desorption from a surface (see Supplementary Information for the theoretical basis). Understanding how related parameters such as slot width, particle composition and particle size affect the release rate of trapped nanoparticles yields information necessary for the engineering of robust, stable slot transport devices.

We carried out a large number of trapping experiments for nanoparticles trapped inside the slot, nanoparticles trapped outside the slot, and the trapped DNA. Trapping was done near the stability point such that the targets would self-release from the waveguide structure. Figure 4c shows a plot of the total number of trapped targets on the waveguide as a function of normalized time. The average release time for particles trapped inside the slot is larger than that for those trapped outside, suggesting greater stability, consistent with our earlier numerical predictions. The trapping stability is related to a kinetic constant,  $k$ , which can be obtained by plotting the above data as a function of reduced time for an appropriate rate law (see Supplementary Information). In Fig. 4d we see that side-trapped particles have a larger rate constant, suggesting that the release (desorption) is faster and a lower work energy is required for release. It is not yet clear why the release process seems to have a second-order rate, but we believe that it results from the exponential decay of the electromagnetic trapping force coupled with hydrodynamic drag. DNA shows a lower stability than the polystyrene nanoparticles, probably because of the extended conformation that it obtains during trapping. As the trapping stability is likely to be strongly dependent on molecular conformation, analysis of the release kinetics in such systems may result in a new method of single-molecule analysis.

Sub-wavelength slot waveguides such as those used in this work can be integrated into lab-on-a-chip platforms using existing manufacturing techniques. They allow discrete optical manipulation and transport of nanoscopic objects with greater precision than is available with existing approaches. The fusion of nanofluidics and optical manipulation in this manner could lead to new methods of bioanalysis and directed assembly.

## METHODS SUMMARY

We fabricate the slot waveguide chips using electron beam lithography. After experimental use, the chips are cleaned using Nanostrip, and scum is removed by using a reactive ion oxygen plasma etching process. The total width of the waveguides is 450 nm with slot widths ranging from 60 nm to 120 nm. The slot waveguides are transitioned to nanotaper devices clad in silicon oxide to increase the coupling efficiency. The laser source is a tunable 1,550-nm laser that runs to a tapered lensed fibre, using the same set-up previously used for our SU-8 waveguide experiments<sup>20</sup>.

The particle solution consists of suspended fluorescent polystyrene nanoparticles 75 nm and 100 nm in diameter (Duke Scientific) with refractive index  $n = 1.574$  in a 100 mM phosphate buffer solution. The particles have a dispersity in diameter of about 10%. The high ionic concentration of the buffer solution suppresses electrostatic interactions in the system and maintains a constant pH during experiments. We add 1%-by-volume Triton X-100 non-ionic surfactant to the particle solution to prevent aggregation of the nanoparticles and to limit adhesion of particles to the surface of the devices and PDMS microchannels.

The experiments use devices that are bonded to a PDMS microchannel 100  $\mu\text{m}$  wide and 5  $\mu\text{m}$  tall. The fluidics are driven using an adjustable air-pressure system designed to maintain a constant pressure to the device. The power output of the fibre during the trapping experiments was set from 250 to 300 mW. Particle trapping was confirmed by counting immobilized particles and counting the number of released particles. We determined particle velocity measurements and particle trapping times by using the ImageJ particle tracking software. Images of the experiments were captured at a rate of 55 ms per frame using a SensiCam CCD camera.

As mentioned in the main text, we carried out experiments demonstrating  $\lambda$ -DNA trapping using the technique described above, with the exception of the smaller 60-nm-wide slot waveguide. The  $\lambda$ -DNA molecules (New England BioLabs) were stained with YOYO-1 intercalating dye (Molecular Probes) so that they could be observed using traditional fluorescence microscopy. The buffer consisted of 10 mM Tris Base (J.T. Baker), 1 mM EDTA (Fisher), and 10 mM sodium chloride (Mallinckrodt) at a pH of 7.8. Poly( $n$ -vinylpyrrolidone) (PVP, Sigma) 2% (by weight) was added to reduce unspecific binding of DNA to channel walls.

Received 31 May; accepted 21 October 2008.

- Grier, D. G. A revolution in optical manipulation. *Nature* **424**, 810–816 (2003).
- Ashkin, A., Dziedzic, J. M., Bjorkholm, J. E. & Chu, S. Observation of a single-beam gradient force optical trap for dielectric particles. *Opt. Lett.* **11**, 288–290 (1986).
- MacDonald, M. P., Spalding, G. C. & Dholakia, K. Microfluidic sorting in an optical lattice. *Nature* **426**, 421–424 (2003).
- Grigorenko, A. N., Roberts, N. W., Dickinson, M. R. & Zhang, Y. Nanometric optical tweezers based on nanostructured substrates. *Nature Photon.* **2**, 365–370 (2008).
- Born, M. & Wolf, E. *Principles of Optics* (Pergamon, 2003).
- Almeida, V. R., Xu, Q. F., Barrios, C. A. & Lipson, M. Guiding and confining light in void nanostructure. *Opt. Lett.* **29**, 1209–1211 (2004).
- Chiu, D. T. & Zare, R. N. Biased diffusion, optical trapping, and manipulation of single molecules in solution. *J. Am. Chem. Soc.* **118**, 6512–6513 (1996).
- Hirano, K. *et al.* Sizing of single globular DNA molecules by using a circular acceleration technique with laser trapping. *Anal. Chem.* **80**, 5197–5202 (2008).
- Psaltis, D., Quake, S. R. & Yang, C. H. Developing optofluidic technology through the fusion of microfluidics and optics. *Nature* **442**, 381–386 (2006).
- Monat, C., Domachuk, P. & Eggleton, B. J. Integrated optofluidics: A new river of light. *Nature Photon.* **1**, 106–114 (2007).
- Neale, S. L., MacDonald, M. P., Dholakia, K. & Krauss, T. F. All-optical control of microfluidic components using form birefringence. *Nature Mater.* **4**, 530–533 (2005).
- Chiou, P. Y., Ohta, A. T. & Wu, M. C. Massively parallel manipulation of single cells and microparticles using optical images. *Nature* **436**, 370–372 (2005).
- Liu, G. L., Kim, J., Lu, Y. & Lee, L. P. Optofluidic control using photothermal nanoparticles. *Nature Mater.* **5**, 27–32 (2006).
- Wang, M. M. *et al.* Microfluidic sorting of mammalian cells by optical force switching. *Nature Biotechnol.* **23**, 83–87 (2005).
- Sinclair, G. *et al.* Assembly of 3-dimensional structures using programmable holographic optical tweezers. *Opt. Express* **12**, 5475–5480 (2004).
- Dholakia, K. & Reece, P. J. in *Structured Light and Its Applications* (ed. Andrews, D. L.) 107–138 (Academic, 2008).
- Righini, M., Zelenina, A. S., Girard, C. & Quidant, R. Parallel and selective trapping in a patterned plasmonic landscape. *Nature Phys.* **3**, 477–480 (2007).
- Marchington, R. F. *et al.* Optical deflection and sorting of microparticles in a near-field optical geometry. *Opt. Express* **16**, 3712–3726 (2008).
- Cizmar, T. *et al.* Optical sorting and detection of submicrometer objects in a motional standing wave. *Phys. Rev. B* **74**, 035105 (2006).
- Schmidt, B. S., Yang, A. H., Erickson, D. & Lipson, M. Optofluidic trapping and transport on solid core waveguides within a microfluidic device. *Opt. Express* **15**, 14322–14334 (2007).
- Ng, L. N., Luf, B. J., Zervas, M. N. & Wilkinson, J. S. Forces on a Rayleigh particle in the cover region of a planar waveguide. *J. Lightwave Technol.* **18**, 388–400 (2000).
- Grujic, K., Helleso, O. G., Hole, J. P. & Wilkinson, J. S. Sorting of polystyrene microspheres using a Y-branched optical waveguide. *Opt. Express* **13**, 1–7 (2005).
- Gaugiran, S., Gétin, S., Fedeli, J. M. & Derouard, J. Polarization and particle size dependence of radiative forces on small metallic particles in evanescent optical fields. Evidences for either repulsive or attractive gradient forces. *Opt. Express* **15**, 8146–8156 (2007).
- Ng, L. N., Luf, B. J., Zervas, M. N. & Wilkinson, J. S. Propulsion of gold nanoparticles on optical waveguides. *Opt. Commun.* **208**, 117–124 (2002).
- Gaugiran, S. *et al.* Optical manipulation of microparticles and cells on silicon nitride waveguides. *Opt. Express* **13**, 6956–6963 (2005).



26. Yang, A. H. J. & Erickson, D. Stability analysis of optofluidic transport on solid-core waveguiding structures. *Nanotechnology* **19**, 045704 (2008).
27. Hart, S. J., Terray, A., Leski, T. A., Arnold, J. & Stroud, R. Discovery of a significant optical chromatographic difference between spores of *Bacillus anthracis* and its close relative, *Bacillus thuringiensis*. *Anal. Chem.* **78**, 3221–3225 (2006).
28. Svoboda, K. & Block, S. M. Optical trapping of metallic Rayleigh particles. *Opt. Lett.* **19**, 930–932 (1994).
29. Braslavsky, I., Hebert, B., Kartalov, E. & Quake, S. R. Sequence information can be obtained from single DNA molecules. *Proc. Natl Acad. Sci. USA* **100**, 3960–3964 (2003).
30. Woolley, A. T., Guillemette, C., Cheung, C. L., Housman, D. E. & Lieber, C. M. Direct haplotyping of kilobase-size DNA using carbon nanotube probes. *Nature Biotechnol.* **18**, 760–763 (2000).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank A. Stroock for discussions, and A. Nitkowski and S. Manipatruni for technical support. This work was done in part at the Cornell NanoScale Facility, a member of the National Nanotechnology Infrastructure Network, which is supported by the US National Science Foundation. This work was funded by the US National Science Foundation NIRT: Active Nanophotofluidic Systems for Single Molecule/Particle Analysis (award number 0708599).

**Author Contributions** A.H.J.Y. and S.D.M. were responsible for running the bulk of the trapping experiments and analysing data. A.H.J.Y. carried out the simulation calculations. B.S.S. was responsible for running initial experiments and fabrication of the slot waveguide chips. M.K. developed the DNA imaging methods. A.H.J.Y., M.L. and D.E. were responsible for writing the paper. All authors discussed the results and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to D.E. (de54@cornell.edu).

## LETTERS

# Seismic evidence for overpressured subducted oceanic crust and megathrust fault sealing

Pascal Audet<sup>1,†</sup>, Michael G. Bostock<sup>1</sup>, Nikolas I. Christensen<sup>1,2</sup> & Simon M. Peacock<sup>1</sup>

Water and hydrous minerals play a key part in geodynamic processes at subduction zones<sup>1–3</sup> by weakening the plate boundary, aiding slip and permitting subduction—and indeed plate tectonics—to occur<sup>4</sup>. The seismological signature of water within the forearc mantle wedge is evident in anomalies with low seismic shear velocity marking serpentinization<sup>5–7</sup>. However, seismological observations bearing on the presence of water within the subducting plate itself are less well documented. Here we use converted teleseismic waves to obtain observations of anomalously high Poisson's ratios within the subducted oceanic crust from the Cascadia continental margin to its intersection with forearc mantle. On the basis of pressure, temperature and compositional considerations, the elevated Poisson's ratios indicate that water is pervasively present in fluid form at pore pressures near lithostatic values. Combined with observations of a strong negative velocity contrast at the top of the oceanic crust, our results imply that the megathrust is a low-permeability boundary. The transition from a low- to high-permeability plate interface downdip into the mantle wedge is explained by hydrofracturing of the seal by volume changes across the interface caused by the onset of crustal eclogitization and mantle serpentinization. These results may have important implications for our understanding of seismogenesis, subduction zone structure and the mechanism of episodic tremor and slip.

In recent years, it has become increasingly apparent that large quantities of water (H<sub>2</sub>O) are stored in the mantle wedge as the serpentine mineral antigorite<sup>5</sup>. Compelling evidence for the presence of antigorite comes from seismic observations that reveal the presence of highly resolved, low-shear-velocity anomalies in both warm (young slab, for example Cascadia<sup>6</sup>) and cold (old slab, for example northeast Japan<sup>7</sup>) subduction zone environments in locations predicted by thermal and petrological models<sup>5</sup>. Water in the mantle wedge most probably originates from the dehydration of hydrous minerals within subducting oceanic crust and mantle, which will occur at shallower depths (<50 km) in warmer environments<sup>1,3</sup>. Hydrous minerals form within oceanic lithosphere (1–2% H<sub>2</sub>O in the crust<sup>1</sup>) as a result of hydrothermal circulation and alteration at spreading ridges, along fracture zones, and in the region of the trench and outer rise. There is little constraint, however, on the *in-situ* abundance and distribution of H<sub>2</sub>O in the fluid and hydrous minerals within subducting oceanic crust, even though this knowledge is likely to be important in assessing the strength of the plate interface and nature of seismogenesis. Recent studies of episodic tremor and slip (ETS) in both Cascadia and Japan point to an origin that involves, either directly or indirectly, fluids near the plate boundary, downdip of the locked zone but updip of its intersection with forearc mantle<sup>8–10</sup>. Characterizing the physical state of subducted oceanic crust and distribution of fluids therein may help in understanding the temporal and spatial occurrence of ETS events.

We use observations of converted teleseismic waves (or receiver functions) recorded over a POLARIS portable array of seismometers across Vancouver Island at spacings of approximately 10 km (Fig. 1). All receiver functions from this experiment are plotted in raw form from west-southwest to east-northeast according to station position and, for individual stations, according to their sampling points (Fig. 2a). These data are sensitive to structures with scale lengths of 1–10 km and are dominated across the entire profile by the signature of an eastward-dipping low-velocity zone. This signature includes three sets of oppositely polarized pulses comprising forward-scattered P-to-S (Ps, 3–4 s at station PGC) conversions, and back-scattered P-to-S (Pps, 13–15 s at PGC) and S-to-S (Pss, 17–20 s at PGC) conversions afforded by reflection of the teleseismic wavefield at the Earth's free surface. These signals coincide with prominent, dipping seismic reflectors and have been previously interpreted to represent oceanic crust of the subducting Juan de Fuca plate<sup>11</sup>, consistent with its expression in studies further south beneath Oregon<sup>5</sup> and worldwide<sup>6,12,13</sup>.

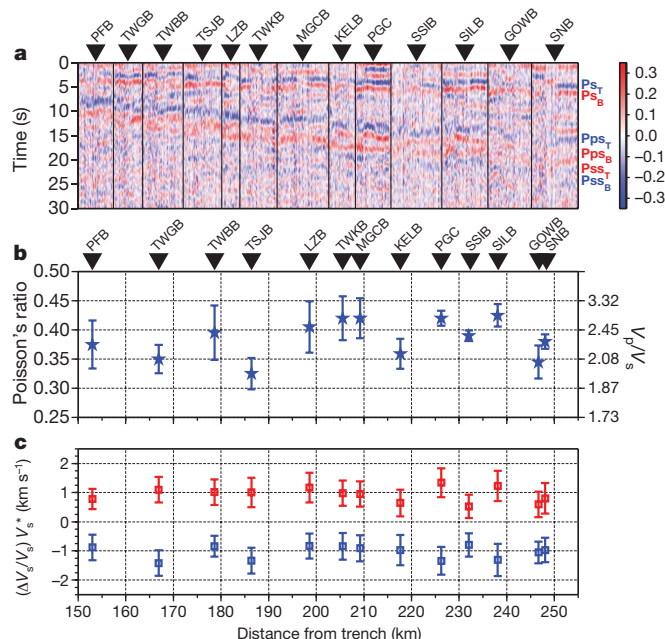
Here we re-examine the signature of the different scattered modes in more detail. It is widely appreciated<sup>14,15</sup> that the timing (relative to the incident P-wave arrival, time 0 in Fig. 2a) of the scattered modes from a given interface can be used to constrain both depth and average



**Figure 1 | Geographical map of northern Cascadia subduction zone.** Location of the three-component, broadband seismic stations that recorded data used in this study are shown as inverted red triangles. The deformation front is indicated by the thin solid black line. Note that the array is oriented approximately perpendicular to the strike of subduction, which trends N45°W.

<sup>1</sup>Department of Earth and Ocean Sciences, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. <sup>2</sup>Department of Geology and Geophysics, University of Wisconsin-Madison, 1215 West Dayton Street, Madison, Wisconsin 53706, USA. †Present address: Seismological Laboratory, University of California Berkeley, 215 McCone Hall, Berkeley, California 94720, USA.





**Figure 2 | Radial receiver function results beneath northern Cascadia.**

**a**, Receiver function results for all data filtered between 0.05 and 0.5 Hz used in the analysis sorted by station position along the array and, for each station, by back-azimuth of incident wavefield. Amplitudes are relative to incident P-wave. Red and blue colours correspond to velocity increase and decrease with depth, respectively. Forward- (Ps) and back-scattered (Pps and Pss) phases show oppositely polarized pulses that are manifest of waves converted from the top and bottom boundaries of a dipping, low-velocity layer, indicated by blue and red labels. **b**, Estimated  $V_P/V_S$  and Poisson's ratios for the low-velocity layer. Errors are standard deviations from bootstrap estimates. Elevated values ( $\sim 0.4$ ) are interpreted as near-lithostatic pore-fluid pressures. **c**, S-velocity perturbations relative to a one-dimensional reference model ( $V_S^*$ ) using a Born approximation for radial Ps phases. Error bars are one standard deviation. Positive (red) and negative (blue) velocity contrasts from the bottom and top, respectively, of the low-velocity layer are of similar amplitude.

P-to-S velocity ratio ( $V_P/V_S$ ) of the overlying column. It can also be shown (see Supplementary Information) that relative times between two arrivals from distinct boundaries within the lithospheric column for each set of scattered modes can be used similarly to characterize the interval properties between the two boundaries, largely independent of overlying structure, even for dipping layers.

This recognition allows us to investigate the velocity structure (that is,  $V_P/V_S$ ) of the downgoing oceanic crust. The key constraint is the ratio of the direct conversion delay time ( $\Delta T_{Ps}$ , the difference in time between direct conversions from the top and bottom of the oceanic crust) to the corresponding back-scattered quantities ( $\Delta T_{Pps}$ ,  $\Delta T_{Pss}$ ). In the two extremes as  $V_P/V_S$  tends to 1 and  $\infty$ ,  $\Delta T_{Ps}$  will tend to 0 and  $\Delta T_{Pps}$  ( $\Delta T_{Pss}/2$ ), respectively. Care must be taken in measuring these quantities, because low-pass filtering at corner periods greater than twice the true separation between pulses will bias measurements of  $\Delta T_{Ps}$ ,  $\Delta T_{Pps}$  and  $\Delta T_{Pss}$  to greater values (see Supplementary Information).

Figure 2b shows our estimates of oceanic crustal  $V_P/V_S$  beneath each station of the array. These estimates are extremely high and it becomes more practical to speak in terms of Poisson's ratio ( $\sigma$ ), for which there is an upper bound of 0.5 for inviscid fluids) around 0.4. These values cannot be due to compositional effects (that is, mineralogy within the downgoing oceanic crust). The common oceanic crustal rocks basalt, diabase and gabbro have  $\sigma$  between 0.28 and 0.29. At the prevailing temperatures and pressures ( $\sim 1$  GPa,  $\sim 500$  C), these rocks will have been metamorphosed to upper amphibolite facies leading to a drop in  $\sigma$  to 0.26 or 0.27 (ref. 16). Antigorite, the high-temperature variety of serpentine, has a  $\sigma$  of 0.29 (ref. 17).

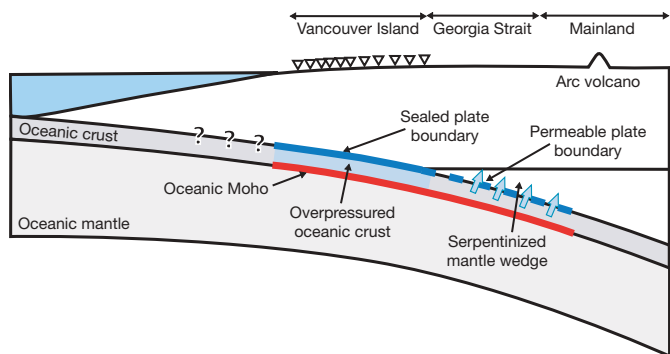
Even the most extreme candidates, the low-pressure serpentine minerals chrysotile and lizardite<sup>16</sup>, have  $\sigma$  below 0.37, and their presence can be further ruled out on the basis of pressure, temperature and compositional considerations<sup>18</sup>. The only plausible explanation is that of high pore-fluid pressures. Widely documented high pore-fluid pressures existing at shallower levels in accretionary prisms are due primarily to porosity reduction<sup>19</sup> but it is more likely that our observations indicate mineral dehydration<sup>20</sup>. Unfortunately, a precise correspondence between  $\sigma$  and pore-fluid pressure is hampered by the lack of laboratory measurements at confining pressures corresponding to 20–40 km depth. Nonetheless, extrapolation of measurements made at lesser pressures<sup>20</sup> shows that it is likely that the oceanic crust beneath Vancouver Island is overpressured (near lithostatic), a result with a range of important implications.

Evidence for high pore-fluid pressure in the Nankai subduction zone has been presented recently<sup>9,21</sup> but these studies, using different forms of travel-time tomography, have found less extreme values of  $V_P/V_S$ . Measurements of  $V_P/V_S$  derived from tomography are likely to be biased towards normal values through the blurring effects of regularization, simultaneous earthquake relocation and the inherent difficulties in imaging low-velocity zones with travel times<sup>22,23</sup>. Estimates of  $\sigma$  obtained using scattered waves are likely to be better resolved owing to their greater sensitivity to local velocity changes.

Our documentation of high pore-fluid pressure implies that the plate interface probably represents a low-permeability boundary. Using Darcy's law, we estimate the permeability of the interface to be  $\sim 5 \times 10^{-25}$  to  $\sim 5 \times 10^{-22}$  m<sup>2</sup>, assuming pore-fluid pressures equal to lithostatic pressures beneath the interface, a 0.65-GPa pressure drop across the interface corresponding to hydrostatic pore-fluid pressure above the interface at a depth of 35 km, a slab fluid production rate of  $10^{-4}$  m<sup>3</sup> (m<sup>2</sup> yr)<sup>-1</sup> (ref. 5), an interface thickness of 1 to 1,000 m, and a dynamic viscosity of  $10^{-4}$  Pa s for H<sub>2</sub>O at 500 C and 1 GPa. These estimates are lower than measurements of basaltic oceanic basement permeabilities in the range  $10^{-21}$  to  $10^{-11}$  m<sup>2</sup> made at length scales of 1 to 1,000 m (ref. 24). The significance of our estimate is reinforced by noting that the greatest permeabilities are found within the upper few hundred metres of oceanic crust and increase with scale length<sup>24</sup>. The low permeability of the plate interface may reflect deformation-induced grain-size reduction<sup>25</sup> or the precipitation of minerals from migrating fluids<sup>26,27</sup>.

Further evidence supporting a low-permeability plate boundary at crustal levels stems from scattered wave amplitudes that are sensitive to velocity contrasts at interfaces (Fig. 2c). Amplitudes of arrivals for a given scattering mode originating from the bottom and top oceanic crustal boundaries are comparable, indicating similar S-velocity jumps of  $\pm 1$  km s<sup>-1</sup>, respectively. The lower boundary jump is consistent with that expected for the oceanic Moho which marks the juxtaposition of lower-velocity gabbro with higher-velocity mantle peridotite. The lower crust of the overriding plate is thought to be formed of dominantly mafic accreted terranes<sup>28</sup> and would thus be expected to afford little or no contrast in seismic velocity with oceanic crust. Prominent amplitudes of scattered waves from the upper boundary thus support a sharp contrast from low-velocity oceanic crust containing trapped fluids to higher velocity overriding crust characterized by much lower pore-fluid pressure.

Low permeability of the plate boundary at depths less than 35 km contrasts with the seismic expression of forearc mantle serpentinization, which appears downdip at depths greater than 40 km in this region<sup>5,11</sup> and suggests that H<sub>2</sub>O is readily transported into the wedge (Fig. 3). A change in the nature of the plate boundary seal can be explained by the large ( $>10\%$ ) volume reduction and release of H<sub>2</sub>O that accompanies eclogitization<sup>29</sup>, the last significant metamorphic facies transition experienced by oceanic crust. Both this reaction and the still greater volume expansion accompanying serpentinization of the mantle wedge<sup>30</sup> could be expected to increase permeability near the plate boundary through fracture generation. This interpretation is consistent with the seismic signature in the cold subduction zone



**Figure 3 | Schematic interpretation of receiver function results.** Updip limit of overpressured oceanic crust is unconstrained. Note vertical exaggeration.

setting of northeast Japan, where a serpentinite layer is interpreted to form above subducting oceanic crust at onset of eclogitization near ~90 km depth<sup>6</sup>.

It is worth noting that in warm subduction zones, ETS tends to occur in the vicinity of the plate interface near but seaward of the wedge corner<sup>8,9</sup>. Our documentation of high pore-fluid pressures in this region supports two groups of models that involve fluids: (1) ETS is triggered by hydrofracturing at the plate interface<sup>10</sup>; and (2) ETS occurs where high pore-fluid pressures extend the region of conditionally stable slip<sup>9,21</sup>. A downdip change in permeability across the plate interface could be explained by hydrofracturing of the seal where hydraulic conductivities become sufficiently high near the onset of eclogitization. Moreover, it might represent an important factor in altering the mode of slip along the thrust within the frictional stability transition zone. Our model suggests the possibility of linking the recurrence of ETS at regular time intervals to periodic cycles of steady pore-fluid pressure build-up from dehydration of subducted oceanic crust, fluid release from fracturing of the interface during ETS, and subsequent precipitation sealing of the plate boundary.

Finally, our documentation of high pore-fluid pressure in the subducting plate has important implications for the interpretation of tomographic velocity images, and the position of the plate boundary interface. High pore-fluid pressures result not only in elevated  $\sigma$  (and  $V_P/V_S$ ) but also in depressed absolute velocities (–20% and –30% for P-waves and S-waves, respectively, in the measurements of Christensen<sup>16</sup>). Consequently, identification of a tomographic iso-velocity contour that does not account for this effect, as proxy for the oceanic Moho<sup>28</sup>, will yield interface depths that are significantly overestimated<sup>11</sup>.

Received 20 May; accepted 17 November 2008.

1. Peacock, S. M. Fluid processes in subduction zones. *Science* **248**, 329–337 (1990).
2. Stern, R. J. Subduction zones. *Rev. Geophys.* **40**, doi:10.1029/2001RG000108 (2002).
3. Kirby, S. H., Engdahl, E. R. & Denlinger, R. in *Subduction Top to Bottom* (eds Bebout, G. E., Scholl, D. W., Kirby, S. H. & Platt, J. P.) 195–214 (Am. Geophys. Un., 1996).
4. O'Neill, C., Jellinek, A. M. & Lenardic, A. Conditions for the onset of plate tectonics on terrestrial planets and moons. *Earth Planet. Sci. Lett.* **261**, 20–32 (2007).
5. Hyndman, R. D. & Peacock, S. M. Serpentinization of the forearc mantle. *Earth Planet. Sci. Lett.* **212**, 417–432 (2003).

6. Bostock, M. G., Hyndman, R. S., Rondenay, S. & Peacock, S. M. An inverted continental Moho and serpentinization of the forearc mantle. *Nature* **417**, 536–539 (2002).
7. Kawakatsu, H. & Watada, S. Seismic evidence for deep-water transportation in the mantle. *Science* **316**, 1468–1471 (2007).
8. Kao, H. et al. A wide depth distribution of seismic tremors along the northern Cascadia margin. *Nature* **436**, 841–844 (2005).
9. Shelly, D. R., Beroza, G. C. & Ide, S. Low-frequency earthquakes in Shikoku, Japan, and their relationship to episodic tremor and slip. *Nature* **442**, 488–491 (2006).
10. Wang, Z., Zhao, D., Mishra, O. P. & Yamada, A. Structural heterogeneity and its implications for the low frequency tremors in Southwest Japan. *Earth Planet. Sci. Lett.* **251**, 66–78 (2006).
11. Nicholson, T., Bostock, M. G. & Cassidy, J. F. New constraints on subduction zone structure in northern Cascadia. *Geophys. J. Int.* **161**, 849–859 (2005).
12. Yuan, X. et al. Subduction and collision processes in the central Andes constrained by converted seismic phases. *Nature* **408**, 958–961 (2000).
13. Abers, G. A., van Keken, P. E., Kneller, E. A., Ferris, A. & Stachnik, J. C. The thermal structure of subduction zones constrained by seismic imaging: Implications for slab dehydration and wedge flow. *Earth Planet. Sci. Lett.* **241**, 387–397 (2006).
14. Zandt, G. & Ammon, C. J. Continental crust composition constrained by measurements of crustal Poisson's ratio. *Nature* **374**, 152–154 (1995).
15. Zhu, L. & Kanamori, H. Moho depth variation in southern California from teleseismic receiver functions. *J. Geophys. Res.* **105**, 2969–2980 (2000).
16. Christensen, N. I. Poisson's ratio and crustal seismology. *J. Geophys. Res.* **101**, 3139–3156 (1996).
17. Christensen, N. I. Serpentinites, peridotites, and seismology. *Int. Geol. Rev.* **46**, 795–816 (2004).
18. Rondenay, S., Abers, G. A. & van Keken, P. E. Seismic imaging of subduction zone metamorphism. *Geology* **36**, 275–278 (2008).
19. Moore, J. C. & Vrolijk, P. Fluids in accretionary prisms. *Rev. Geophys.* **30**, 113–135 (1992).
20. Christensen, N. I. Pore pressure and oceanic crustal seismic structure. *Geophys. J. R. Astron. Soc.* **79**, 411–423 (1984).
21. Kodaira, S. et al. High pore fluid pressure may cause silent slip in the Nankai Trough. *Science* **304**, 1295–1298 (2004).
22. Gerver, M. & Markushkevitch, V. Determination of a seismic wave velocity from the travel time curve. *Geophys. J. R. Astron. Soc.* **11**, 165–173 (1966).
23. Wielandt, E. in *Seismic Tomography* (ed. Nolet, G.) 85–98 (Reidel, 1987).
24. Fisher, A. T. Permeability within basaltic oceanic crust. *Rev. Geophys.* **36**, 143–182 (1998).
25. Caine, J. S., Evans, J. P. & Foster, C. B. Fault zone architecture and permeability structure. *Geology* **24**, 1025–1028 (1996).
26. Kato, A., Sakaguchi, A., Yoshida, S. & Mochizuki, H. Permeability measurements and precipitation sealing of basalt in an ancient exhumed fault of a subduction zone. *Bull. Earthq. Res. Inst. Univ. Tokyo* **78**, 83–89 (2003).
27. Meneghini, F. & Moore, J. C. Deformation and hydrofracturing in a subduction thrust at seismogenic depths: The Rodeo Cove thrust zone, Marin Headlands, California. *Geol. Soc. Am. Bull.* **119**, 174–183 (2007).
28. Ramachandran, K., Dosso, S. E., Spence, G. D., Hyndman, R. D. & Brocher, T. M. Forearc structure beneath southwestern British Columbia: A three-dimensional tomographic velocity model. *J. Geophys. Res.* **110**, B02303 (2005).
29. Ahrens, T. J. & Schubert, G. Gabbro–eclogite reaction rate and its geophysical significance. *Rev. Geophys.* **13**, 383–400 (1975).
30. Coleman, R. G. Petrologic and geophysical nature of serpentinites. *Geol. Soc. Am. Bull.* **82**, 897–918 (1971).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank E. Davis and M. Jellinek for discussions. Data used in this study come from the Canadian National Seismological Network and are distributed freely by the Geological Survey of Canada.

**Author Contributions** P.A. and M.G.B. designed the study, analysed the data and wrote the paper; N.I.C. and S.M.P. participated in the interpretation of results.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to P.A. ([paudet@berkeley.edu](mailto:paudet@berkeley.edu)).



# Indirect reciprocity provides only a narrow margin of efficiency for costly punishment

Hisashi Ohtsuki<sup>1,2</sup>, Yoh Iwasa<sup>3</sup> & Martin A. Nowak<sup>4</sup>

Indirect reciprocity<sup>1–5</sup> is a key mechanism for the evolution of human cooperation. Our behaviour towards other people depends not only on what they have done to us but also on what they have done to others. Indirect reciprocity works through reputation<sup>5–17</sup>. The standard model of indirect reciprocity offers a binary choice: people can either cooperate or defect. Cooperation implies a cost for the donor and a benefit for the recipient. Defection has no cost and yields no benefit. Currently there is considerable interest in studying the effect of costly (or altruistic) punishment on human behaviour<sup>18–25</sup>. Punishment implies a cost for the punished person. Costly punishment means that the punisher also pays a cost. It has been suggested that costly punishment between individuals can promote cooperation. Here we study the role of costly punishment in an explicit model of indirect reciprocity. We analyse all social norms, which depend on the action of the donor and the reputation of the recipient. We allow errors in assigning reputation and study gossip as a mechanism for establishing coherence. We characterize all strategies that allow the evolutionary stability of cooperation. Some of those strategies use costly punishment; others do not. We find that punishment strategies typically reduce the average payoff of the population. Consequently, there is only a small parameter region where costly punishment leads to an efficient equilibrium. In most cases the population does better by not using costly punishment. The efficient strategy for indirect reciprocity is to withhold help for defectors rather than punishing them.

Human societies are organized around cooperative interactions. But why would natural selection equip selfish individuals with altruistic tendencies? This question has fascinated evolutionary biologists for decades. One answer is given in terms of direct reciprocity<sup>26–29</sup>. There are repeated encounters between the same two individuals: I help you, and you help me. More recently, indirect reciprocity has emerged as a more general model: I help you, and somebody helps me. Indirect reciprocity is based on reputation<sup>5</sup>. People monitor the social interactions within their group. Helping others establishes the reputation of being a helpful individual. Natural selection can favour strategies that help those who have helped others<sup>5–17</sup>. The consequences for widespread cooperation are enormous. Direct reciprocity is like an economy based on the exchange of goods, whereas indirect reciprocity resembles the invention of money. The money that feeds the engines of indirect reciprocity is reputation. For direct reciprocity, my strategy depends on what you have done to me; for indirect reciprocity, my strategy also depends on what you have done to others. Direct and indirect reciprocity are mechanisms for the evolution of cooperation<sup>30</sup>.

Punishment refers to an action that implies a cost for the punished person. Costly punishment means that the punisher also pays a cost for exercising punishment. In certain experimental situations costly punishment has been called ‘altruistic punishment’, because the

punishers cannot expect any material gain from their action<sup>20,21</sup>. In reality, however, most punishment actions among humans are associated with the expectation of a delayed material gain; they are therefore not altruistic.

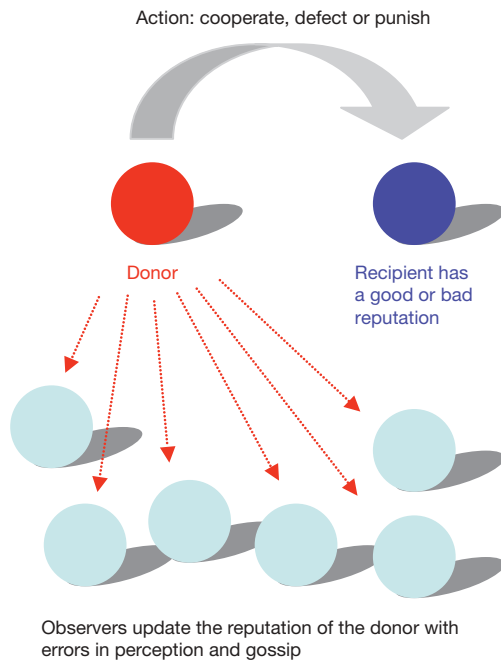
The suggested idea for the evolution of cooperation is that people might be more willing to cooperate under the threat of punishment. However, we note that costly punishment is not a separate mechanism for the evolution of cooperation but a form of direct or indirect reciprocity. If I punish you because you have defected with me, then I use direct reciprocity. If I punish you because you have defected with others, then indirect reciprocity is at work. In the setting of direct reciprocity, punishment is a form of retaliation<sup>25</sup>. For indirect reciprocity, punishment works through reputation and also includes third-party actions, which means that observers of an interaction are willing to punish defectors at a cost to themselves<sup>21</sup>. Therefore, any discussion of the evolution of costly punishment brings us immediately into the framework of direct or indirect reciprocity.

In general, the reputation score could be a continuous variable<sup>5</sup>, but here we consider a simple model with binary reputation. People have either a good reputation (G) or a bad reputation (B). At times, two random players are chosen from the population, one in the role of donor, the other in the role of recipient. The donor can either cooperate (C), defect (D) or punish (P). Cooperation means the donor pays a cost  $c$ , and the recipient gets a benefit  $b$ . Punishment implies that the donor pays a cost  $\alpha$  and the recipient incurs a cost  $\beta$ . For defection there is no cost and no benefit.

The interaction between the donor and the recipient is observed by the other members of the population (Fig. 1). The reputation of the donor is updated according to a social norm. First-order assessment depends only on the action of the donor; for example, cooperation leads to a good reputation, whereas defection leads to a bad reputation. Second-order assessment<sup>12,13</sup> depends both on the action of the donor and the reputation of the recipient: for example, it could be deemed ‘good’ to cooperate with a good recipient but ‘bad’ to cooperate with a bad recipient. Here we study social norms that use second-order assessment. The donor has three possible moves (C, D or P) and the recipient has one of two reputations (G or B). There are therefore six combinations and  $2^6 = 64$  social norms with second-order assessment. All detailed calculations are shown in the Supplementary Information.

Any interaction leads to either a good or a bad reputation for the donor. We assume that this process of reputation updating is subject to errors. There may be wrong observations or the spread of false rumours. With probability  $\mu$  an incorrect reputation is assigned and adopted by all. In the simplest model, everyone has the same opinion of everyone else. There are no private lists of reputation. Triggering a wrong reputation affects everyone equally. The parameter  $q = 1 - 2\mu$  quantifies the ability of the population to distinguish between good

<sup>1</sup>Department of Value and Decision Science, Tokyo Institute of Technology, Tokyo 152-8552, Japan. <sup>2</sup>PRESTO, Japan Science and Technology Agency, 4-1-8 Honcho Kawaguchi, Saitama 332-0012, Japan. <sup>3</sup>Department of Biology, Faculty of Sciences, Kyushu University, Fukuoka 812-8581, Japan. <sup>4</sup>Program for Evolutionary Dynamics, Department of Organismic and Evolutionary Biology, Department of Mathematics, Harvard University, Cambridge, Massachusetts 02138, USA.



**Figure 1 | Indirect reciprocity with costly punishment.** The donor chooses one of three actions: cooperate, defect or punish. The recipient has a binary reputation, which is either 'good' or 'bad'. The donor's choice depends on the recipient's reputation. The donor's action is observed by other members of the population, who update the donor's reputation according to a social norm, which is shared by all. The donor is assigned an incorrect reputation with probability  $\mu$ . The 'social resolution',  $q = 1 - 2\mu$ , is a key parameter of indirect reciprocity: it defines the ability to distinguish between good and bad.

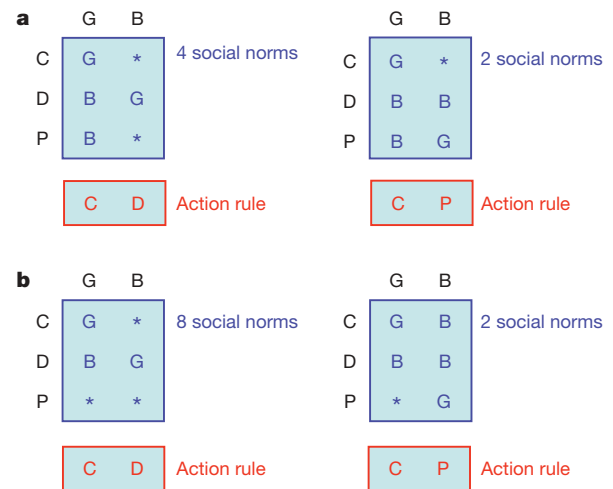
and bad. We call  $q$  the 'social resolution'. If  $\mu = 1/2$  then reputation is assigned at random, and there is no ability to distinguish between good and bad ( $q = 0$ ).

Games of indirect reciprocity contain social norms and action rules. The action rule specifies for the donor whether to cooperate, defect or punish a recipient who is either good or bad. For example, the action rule CD prescribes cooperation with a good recipient and defection with a bad recipient; this rule does not use costly punishment. In contrast, the action rule CP prescribes cooperation with good recipients and punishment of bad recipients. The action rules CC, DD and PP encode, respectively, unconditional cooperation, defection and punishment. In total, there are nine possible action rules.

For each of the 64 social norms we study the competition of all nine action rules. We assume that everyone in the population has the same social norm, and we evaluate whether this norm allows the evolutionary stability of action rules that specify cooperation with good recipients. There are only two candidates for such action rules, CD and CP, because CC is not stable against invasion by defectors (DD). Figure 2 shows all social norms that allow the evolution of cooperation. The action rule DD is evolutionarily stable for any social norm.

Social norms that stabilize the CD action rule have the following properties: (1) cooperation with a good recipient leads to a good reputation; (2) defection against a good recipient leads to a bad reputation; and (3) defection against a bad recipient leads to a good reputation. The three remaining positions in the norm can be either G or B. If the cost of cooperation is greater than the cost of punishment ( $c > \alpha$ ), then punishing a good recipient must lead to a bad reputation; otherwise a donor can keep a good reputation by using the cheaper punishment option instead of the more expensive cooperation move.

Social norms that stabilize the CP action rule have the following properties: first, cooperation with a good recipient leads to a good



**Figure 2 | Social norms of cooperation.** We have determined all social norms that allow the evolutionary stability of action rules prescribing cooperation with good recipients. There are two such action rules, CD and CP. The former 'punishes' bad recipients by defection (D); the latter uses costly punishment (P). In **a** the cost of cooperation exceeds the cost of punishment; in **b** the cost of cooperation is less than the cost of punishment. There is an intuitive summary of all successful social norms: following the action rule maintains a good reputation; deviating from it can lead to a good or bad reputation; a deviation that is less costly than the prescribed action must lead to a bad reputation. An asterisk denotes G or B. The detailed parameter requirements for evolutionary stability are given in the text and in the Supplementary Information.

reputation; second, defection always leads to a bad reputation; and third, punishing a bad recipient leads to a good reputation.

CD action rules are evolutionarily stable, if the social resolution exceeds the cost-to-benefit ratio ( $q > c/b$ ). In contrast, CP action rules are evolutionarily stable if  $q > \max\{c, \alpha\}/(b + \beta)$ . Note that costly punishment can stabilize cooperation even if  $q < c/b$ . Thus, costly punishment can in principle extend the stability range of cooperation. DD action rules are always evolutionarily stable.

We have performed computer simulations in heterogeneous populations of finite size to test the validity of our analytical calculations. We find that the CD and CP action rules are stable against invasion by other action rules under the appropriate social norms and given the right parameter region. In the simplest simulations everyone has the same information about the reputation of others. In the extended simulations we drop this assumption. Now there are individual errors in assigning reputation. Consequently everyone has a private list of the reputation of others. These errors can destroy indirect reciprocity unless there is a mechanism for re-establishing coherence. Gossip is such a mechanism. We assume that individuals talk to each other and sample each other's opinions (as in a 'voter model'). If there are enough communication events, then we observe the evolutionary stability of our strategies as predicted. We have also studied errors resulting in execution of the wrong action ('trembling hand') or recalling an incorrect reputation ('fuzzy mind'). Our results are robust as long as these errors are not too frequent. All simulations are described in the Supplementary Information.

For some parameter regions, multiple action rules are evolutionarily stable. We therefore ask the following question: for all possible parameter regions, which of the three action rules CD, CP and DD are stable, and which one is the most efficient in the sense of leading to the highest average payoff at equilibrium? We obtain the following answer. (1) If  $q > c/b$ , then CD is most efficient. (2) If  $c/b > q > c/(b + \beta)$  then CP is stable and more efficient than DD, if the following two conditions hold:

$$q > \frac{\alpha}{b + \beta} \quad \text{and} \quad q > \frac{\alpha + \beta - b + c}{\alpha + \beta + b - c} \quad (1)$$



Otherwise DD is more efficient than CP. If  $b < c$ , then DD is always more efficient than CP. (3) If  $c/(b + \beta) > q$ , then only DD is evolutionarily stable.

Thus, if the accuracy of assigning the correct reputation,  $q$ , is too low, then only DD is efficient. If  $q$  is sufficiently large, then CD is efficient. For intermediate values of  $q$  there can be a small window where CP is efficient. However, the existence of this parameter region depends on whether the key parameters  $b$ ,  $c$ ,  $\alpha$  and  $\beta$  fulfil the constraints given by equation (1). Let us consider a numerical example. If  $b = 2$ ,  $c = 1$ ,  $\alpha = 1/2$  and  $\beta = 2$ , then CD is efficient for  $q > 1/2$ , whereas CP is efficient for  $1/2 > q > 3/7$  and DD is efficient for  $3/7 > q$ . If we increase the effect of punishment to  $\beta = 5/2$  (or larger), there is no region left where CP is efficient. Intuitively, if CD is evolutionarily stable, it is always the most efficient equilibrium. If it is not stable, the remaining parameter region where CP is stable and more efficient than DD is very small or non-existent. Figure 3 illustrates the narrow margin of efficiency of costly punishment.

These considerations of efficiency do not imply that all populations will evolve towards punishment-free action rules. It is possible that a population is stuck at an inefficient equilibrium for a long time. A model with contingent movement allows us to study the competition

of different social norms. We examine a simple model in which two groups have two different social norms. One norm stabilizes CD, whereas the other norm stabilizes CP. People interact only within their own group, but sometimes they compare their payoff with individuals from the other group. People might move to the other group and adopt its social norm if they find that its members have a higher payoff. We observe rapid selection of the efficient equilibrium (see Supplementary Information).

In an experimental study, the observers of a Prisoner's Dilemma game between two other people sometimes punish defectors at a cost to themselves<sup>21</sup>. This behaviour is a form of indirect reciprocity. In another experiment<sup>23</sup>, a public goods game is followed by one round of punishment and then by one round of cooperation or defection. This setup is not directly comparable with our model, but the observation is that adding the third round reduces the amount of punishment that is being used in the second round. This particular finding is in agreement with our result: other possibilities of indirect reciprocity reduce the amount of costly punishment. In the context of our theory it would be important to extend both experiments to permit reputation-building over multiple rounds of interaction and a choice between cooperation, defection and costly punishment in every round. We predict that such an experiment will show that costly punishment is an inefficient behaviour for most parameter regions.

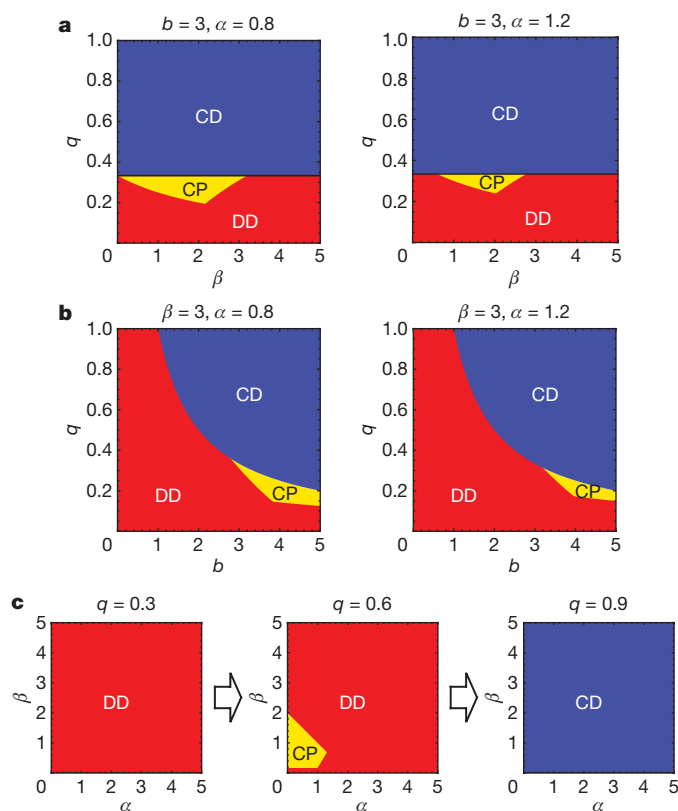
We have studied the effect of costly punishment in an explicit model of indirect reciprocity. We have analysed all social norms that use binary reputation and second-order assessment. We find that both CD and CP action rules can stabilize cooperation. These rules reward good recipients with cooperation and 'punish' bad ones with either defection (CD) or costly punishment (CP). If both CD and CP action rules are evolutionarily stable, the use of costly punishment leads to a lower equilibrium payoff and is therefore inefficient. It is even possible that costly punishment yields a lower payoff than all-out defection (DD). Costly punishment maximizes the group average payoff for only a very limited parameter region. This narrow margin of efficiency requires fine-tuning of the key parameters. If the social resolution exceeds the cost to benefit ratio ( $q > c/b$ ), CD rules are always more efficient than CP rules. The evolution of improved mechanisms of indirect reciprocity therefore leads to societies in which costly punishment between individuals is not an efficient behaviour for promoting cooperation.

## METHODS SUMMARY

An action rule  $s$  is formulated as a mapping from  $\{G, B\}$  (the recipient's reputation) to  $\{C, D, P\}$  (the prescribed action). A social norm  $n$  is a mapping from the product of  $\{C, D, P\}$  (the donor's action) and  $\{G, B\}$  (the recipient's reputation) to  $\{G, B\}$  (the donor's new reputation). We search for the combination of an action rule  $s$  and a social norm  $n$  that satisfies the following two properties: first, the monomorphic population in which all players adopt  $s$  and  $n$  achieves full cooperation in the absence of errors; and second, the action rule  $s$  is evolutionarily stable under the social norm  $n$ . We check these two criteria for each of all  $9 \times 64 = 576$  possible combinations of action rule and social norm ( $s, n$ ). From the first criterion, action rule  $s$  must use cooperation (C). Because of the symmetry in the binary labels G and B we can assume without loss of generality that the action rule prescribes cooperation to good recipients; that is,  $s(G) = C$ . To study the evolutionary stability of the action rule  $s$ , we use dynamic optimization. We assume that the social norm is  $n$  and that all players except the focal player adopt action rule  $s$ . Under this assumption we calculate the best-response action rule  $s^*$  of the focal player. If  $s^*$  exists uniquely and matches  $s$ , then  $s$  is evolutionarily stable under  $n$ . Coexistence of different action rules<sup>10</sup> is not within the scope of our analysis. See the Supplementary Information for further details.

Received 11 June; accepted 3 November 2008.

1. Sugden, R. *The Economics of Rights, Cooperation and Welfare* (Blackwell, 1986).
2. Alexander, R. D. *The Biology of Moral Systems* (Aldine de Gruyter, 1987).
3. Kandori, M. Social norms and community enforcement. *Rev. Econ. Stud.* 59, 63–80 (1992).
4. Okuno-Fujiwara, M. & Postlewaite, A. Social norms and random matching games. *Games Econ. Behav.* 9, 79–109 (1995).
5. Nowak, M. A. & Sigmund, K. Evolution of indirect reciprocity by image scoring. *Nature* 393, 573–577 (1998).



**Figure 3 | The marginal efficiency of costly punishment.** Projections of the five-dimensional ( $b$ ,  $c$ ,  $\alpha$ ,  $\beta$  and  $q$ ) parameter space onto various two-dimensional planes. The parameters  $b$  and  $c$  denote the benefit and cost of cooperation. The parameters  $\alpha$  and  $\beta$  denote the cost and effect of punishment. The social resolution of the system is given by  $q$ , the probability of distinguishing between good and bad in a world where errors in assignment of reputation are possible. The symbols CD, CP and DD represent the region where the corresponding action rule is the most efficient equilibrium, which means the evolutionarily stable strategy with the highest average payoff. CD means cooperation with good recipients and defection with bad ones. CP means cooperation with good recipients and punishment of bad ones. DD is unconditional defection. Costly punishment is an efficient equilibrium only for a very constrained parameter region (shown in yellow). **a**, Projections on the  $b$ - $q$  plane for  $b = 3$  and  $\alpha = 0.8$  or  $1.2$ . **b**, Projections on the  $b$ - $q$  plane for  $\beta = 3$  and  $\alpha = 0.8$  or  $1.2$ . **c**, Projections on the  $\alpha$ - $q$  plane for  $b = 1.5$  and  $q = 0.3, 0.6$  or  $0.9$ . We always use  $c = 1$ .

6. Wedekind, C. & Milinski, M. Cooperation through image scoring in humans. *Science* **288**, 850–852 (2000).
7. Dufwenberg, M., Gneezy, U., Güth, W. & van Damme, E. Direct vs indirect reciprocity: an experiment. *Homo Oecon.* **18**, 19–30 (2001).
8. Fishman, M. A. Indirect reciprocity among imperfect individuals. *J. Theor. Biol.* **225**, 285–292 (2003).
9. Ohtsuki, H. & Iwasa, Y. How should we define goodness?—reputation dynamics in indirect reciprocity. *J. Theor. Biol.* **231**, 107–120 (2004).
10. Brandt, H. & Sigmund, K. The logic of reprobation: assessment and action rules for indirect reciprocation. *J. Theor. Biol.* **213**, 475–486 (2004).
11. Bolton, G. E., Katok, E. & Ockenfels, A. Cooperation among strangers with limited information about reputation. *J. Public Econ.* **89**, 1457–1468 (2005).
12. Brandt, H. & Sigmund, K. Indirect reciprocity, image-scoring, and moral hazard. *Proc. Natl Acad. Sci. USA* **102**, 2666–2670 (2005).
13. Nowak, M. A. & Sigmund, K. Evolution of indirect reciprocity. *Nature* **437**, 1291–1298 (2005).
14. Suzuki, S. & Akiyama, E. Reputation and the evolution of cooperation in sizable groups. *Proc. R. Soc. Lond. B* **272**, 1373–1377 (2005).
15. Chalub, F. A. C. C., Santos, F. C. & Pacheco, J. M. The evolution of norms. *J. Theor. Biol.* **241**, 233–240 (2006).
16. Takahashi, N. & Mashima, R. The importance of subjectivity in perceptual errors on the emergence of indirect reciprocity. *J. Theor. Biol.* **243**, 418–436 (2006).
17. Pacheco, J. M., Santos, F. C. & Chalub, F. A. C. C. Stern-judging: a simple, successful norm which promotes cooperation under indirect reciprocity. *PLoS Comp. Biol.* **2**, 1634–1638 (2006).
18. Yamagishi, T. Seriousness of social dilemmas and the provision of a sanctioning system. *Soc. Psychol. Q.* **51**, 32–42 (1988).
19. Clutton-Brock, T. H. & Parker, G. A. Punishment in animal societies. *Nature* **373**, 209–216 (1995).
20. Fehr, E. & Gächter, S. Altruistic punishment in humans. *Nature* **415**, 137–140 (2002).
21. Fehr, E. & Fischbacher, U. Third-party punishment and social norms. *Evol. Hum. Behav.* **25**, 63–87 (2004).
22. Fowler, J. H. Altruistic punishment and the origin of cooperation. *Proc. Natl Acad. Sci. USA* **102**, 7047–7049 (2005).
23. Rockenbach, B. & Milinski, M. The efficient interaction of indirect reciprocity and costly punishment. *Nature* **444**, 718–723 (2006).
24. Sigmund, K. Punish or perish? Retaliation and collaboration among humans. *Trends Ecol. Evol.* **22**, 593–600 (2007).
25. Dreber, A., Rand, D. G., Fudenberg, D. & Nowak, M. A. Winners don't punish. *Nature* **452**, 348–351 (2008).
26. Trivers, R. L. The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57 (1971).
27. Axelrod, R. & Hamilton, W. D. The evolution of cooperation. *Science* **211**, 1390–1396 (1981).
28. Colman, A. M. *Game Theory and Its Applications in the Social and Biological Sciences* (Routledge, 1995).
29. Rutte, C. & Taborsky, M. The influence of social experience on cooperative behaviour of rats (*Rattus norvegicus*): direct vs generalised reciprocity. *Behav. Ecol. Sociobiol.* **62**, 499–505 (2008).
30. Nowak, M. A. Five rules for the evolution of cooperation. *Science* **314**, 1560–1563 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** Support from the John Templeton Foundation, the Japan Society for the Promotion of Science, the NSF/NIH joint program in mathematical biology (NIH grant R01GM078986) and J. Epstein is gratefully acknowledged.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to H.O. (ohtsuki.h.aa@m.titech.ac.jp).



# Emergence of complex cell properties by learning to generalize in natural scenes

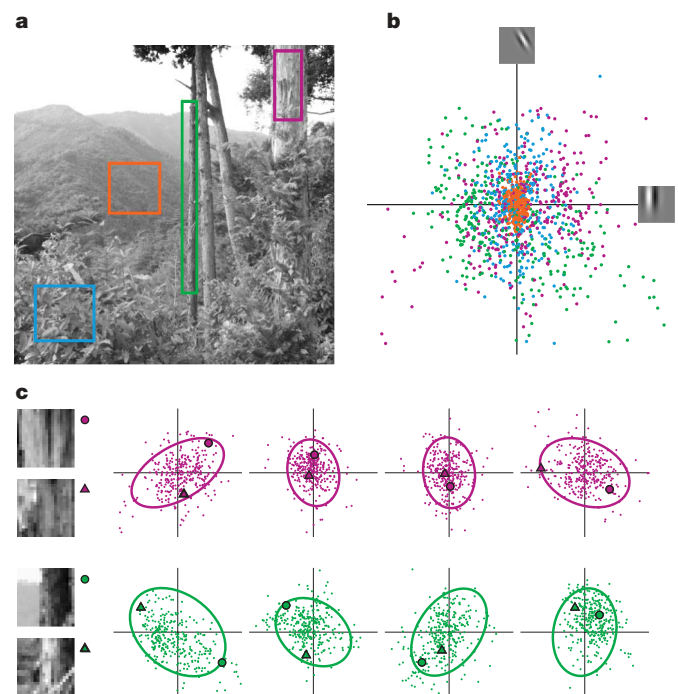
Yan Karklin<sup>1†</sup> & Michael S. Lewicki<sup>1‡</sup>

A fundamental function of the visual system is to encode the building blocks of natural scenes—edges, textures and shapes—that subserve visual tasks such as object recognition and scene understanding. Essential to this process is the formation of abstract representations that generalize from specific instances of visual input. A common view holds that neurons in the early visual system signal conjunctions of image features<sup>1,2</sup>, but how these produce invariant representations is poorly understood. Here we propose that to generalize over similar images, higher-level visual neurons encode statistical variations that characterize local image regions. We present a model in which neural activity encodes the probability distribution most consistent with a given image. Trained on natural images, the model generalizes by learning a compact set of dictionary elements for image distributions typically encountered in natural scenes. Model neurons show a diverse range of properties observed in cortical cells. These results provide a new functional explanation for nonlinear effects in complex cells<sup>3–6</sup> and offer insight into coding strategies in primary visual cortex (V1) and higher visual areas.

As we scan across a complex natural scene, fixations at multiple locations (for example, on the trunk of a tree or along its edge) produce a coherent percept of the underlying structure (the bark texture or the contour of the edge), even though individual images collected at the retina are inherently highly variable. Figure 1 illustrates the problem our brain solves so effortlessly: perceptually distinct image regions produce response patterns that are highly overlapping and cannot be easily distinguished using low-level, linear representations. What sort of computations are required to achieve generalization across natural stimuli?

Early visual neurons are typically described as linear feature detectors<sup>1,2</sup>. Models developed around this idea can accurately capture the behaviour of neurons from the retina<sup>7</sup> to simple cells in the cortex<sup>8</sup> but, as the examples in Fig. 1 illustrate, neither individual features nor linear transformations can reliably discriminate images of one structure from another. More abstract features are presumably computed in later stages of the visual system, but our knowledge of processing by these neurons is limited. In V1, complex cells respond to an edge over a range of positions<sup>1</sup>, but classical models of these cells<sup>9,10</sup> fail to explain a number of nonlinear effects, such as surround suppression and cross-orientation inhibition<sup>3–5</sup>. More importantly, there is no functional explanation for the role of these behaviours in the perception of natural scenes. In higher visual areas such as V2 and V4, neurons are more invariant to image properties such as position and scale<sup>11–13</sup> and might be encoding shape or texture<sup>12,14,15</sup>. For these neurons to generalize effectively, the neural circuitry must generate a representation that is similar across the wide distribution of images of a given type (for example, a texture or contour) yet distinct across the much larger distribution of all other images.

Previous theoretical work has shown that neurons in the primary visual cortex form an efficient code adapted to the statistics of natural images<sup>16,17</sup>, but this says nothing about how neurons generalize across



**Figure 1 | Statistical patterns distinguish local regions of natural scenes.** **a**, A natural scene with four distinct regions outlined (image courtesy of E. Doi). **b**, The scatter plot shows the joint output of a pair of linear feature detectors (oriented Gabor filters) for  $20 \times 20$ -image patches sampled from the four regions. The outputs from different regions are highly overlapping, indicating that linear features provide no means to distinguish between the regions. **c**, Each column shows the joint output of a different pair of linear feature detectors sampled from the regions containing the tree bark or the tree edge (the first column corresponds to features in **b**). The correlations in each panel can be described by a Gaussian distribution and its covariance (ellipses). The differences in the distributions between the rows reveal characteristic patterns in correlations, which become even more prominent as projections onto more features are considered. These patterns can be used to generalize within regions while still distinguishing among them. As an example, we highlight two patches in each region, shown by the circle and triangle in each panel. Although the pairs of images are visibly quite different, each image is consistent with the distribution of the local image region. By contrasting the distributions across multiple dimensions, it is possible to infer image type for single patches, even if the patches have similar projections along some image features.

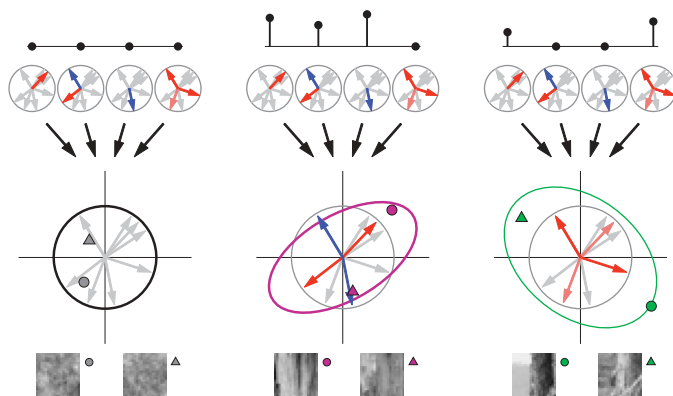
<sup>†</sup>Computer Science Department & Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15213, USA. <sup>‡</sup>Present address: Center for Neural Science, New York University, New York, New York, USA (Y.K.); Electrical Engineering and Computer Science Department, Case Western University, Cleveland, Ohio, USA and Wissenschaftskolleg (Institute for Advanced Study) zu Berlin, Germany (M.S.L.).

the intrinsic variability of scene elements. Here we extend the efficient coding approach and propose that an important aspect of visual computation is to represent the myriad statistical distributions that characterize local image regions. Rather than coding the pixel intensities of a patch of texture or edge, neurons in later stages encode the image distribution (that is, the range and pattern of variability of pixel intensities or image features) that is most consistent with the input image. This allows the neural representation to generalize across individual fixations and convey more abstract properties of the image. We demonstrate that a model designed around this computational goal and optimized for natural scenes explains nonlinear properties of complex cells and neurons in higher visual areas, thereby providing a new functional interpretation for these effects.

Fundamentally, generalization is the identification of common characteristics of a class from specific instances. The goal of the proposed model is to learn the statistical distributions that characterize local image regions, such as those in Fig. 1, and identify them from individual image patches. What statistical regularities are relevant for this task? As the examples in Fig. 1 suggest, the distributions of perceptually similar images show consistent patterns in the degree of variation along some dimensions, as well as in the strength of correlations (and anti-correlations) among different feature dimensions. Although these patterns appear subtle when projected onto two dimensions, as in the examples, the full multivariate distribution, consisting of hundreds of dimensions, produces prominent statistical signatures that can be exploited by the visual system.

To determine how the model generalizes, we must specify how it represents distributions of local image regions. A simple way to summarize the patterns of correlations for a given type of image is the covariance matrix of the data. A neural code for this structure could be defined by enumerating the set of observed covariances and assigning one neuron to each pattern, but this approach presents two problems. First, local image classes are not known a priori. Second, given the limited number of neurons in the visual system, it is not feasible to represent all possible image types, let alone the combinatorial number of possible image boundaries. Instead, we propose a distributed code in which the graded activity of the neural population acts to describe a continuum of potential covariance patterns.

This distribution coding model is illustrated schematically in Fig. 2. The model represents the correlations present in local image regions with a multivariate Gaussian distribution that has a fixed mean of zero and a covariance that is a function of the neural activity (see



**Figure 2 | Distribution coding model.** Rather than encoding the precise pixel values of an input image (bottom), the proposed model infers for each image the most likely distribution (ellipses) containing it. Activation patterns for model neurons are shown at the top of each column. Absence of activity corresponds to the lack of image structure (left panel)—that is, a canonical distribution that reflects the statistics over all natural images (black circle). Increased neural activity represents deviations from this canonical distribution and captures statistical patterns in local image regions (middle and right panels, patches and symbols as in Fig. 1). In each panel, the activation pattern is the same for both inputs. See text for further details.

Methods). This simple statistical description affords both the flexibility to capture a continuum of natural image distributions and mathematical simplicity for tractable parameter estimation. The model uses two sets of parameters to describe correlations in image distributions. First, the vectors  $\mathbf{b}_k$  (arrows within circles) specify image features along which the encoded distribution can be expanded or contracted relative to the canonical distribution (black circle). These vectors are shared by all neurons in the model (represented by the four grey circles, each of which contains the same set of arrows). Because these vectors do not necessarily line up with the axes of the input dimensions, changes in variation along a vector can correspond to changes in the correlational pattern in many dimensions at once. Neurons in the model ( $y_j$ ) describe changes along these directions using weights  $w_{jk}$ : each has a different set of weights, corresponding to an expansion or contraction along feature  $\mathbf{b}_k$ . A positive weight (red) means that the neuron responds to a wider range of stimuli along that direction, a negative weight (blue) means it responds to a narrower range, and a weight close to zero (grey) indicates that the neuron is neutral to this direction. The combined activation of all neurons specifies the final shape of the encoded distribution (ellipses). Given a single fixation—one input image—the model computes the neural representation (that is, the image distribution) that provides the most probable explanation of the input. The model is able to generalize over different image regions if the inferred representation is similar across a region (for example, for the pairs of patches in Fig. 2).

By adapting model parameters  $\mathbf{b}_k$  and  $w_{jk}$  to the data, we are able to find the most efficient way to use a limited number of neurons to describe the wide range of distributions observed in natural images. It should be noted that, although our goal is to derive a stable representation of all patches within a local region, no assumptions about locality are made (encoding is done independently for each image patch). It is the task of the model to learn a compact representation of all patches and to automatically discover which share the statistical properties of a particular type.

If, as hypothesized, neurons in the visual cortex encode patterns in correlations in local regions and are adapted specifically to the statistics of natural scenes, we expect the representations learned by the model to reflect properties of visual neurons. To this end, we trained the model on patches sampled from a large set of natural images and examined the resulting parameters as well as the response properties of model neurons to natural images.

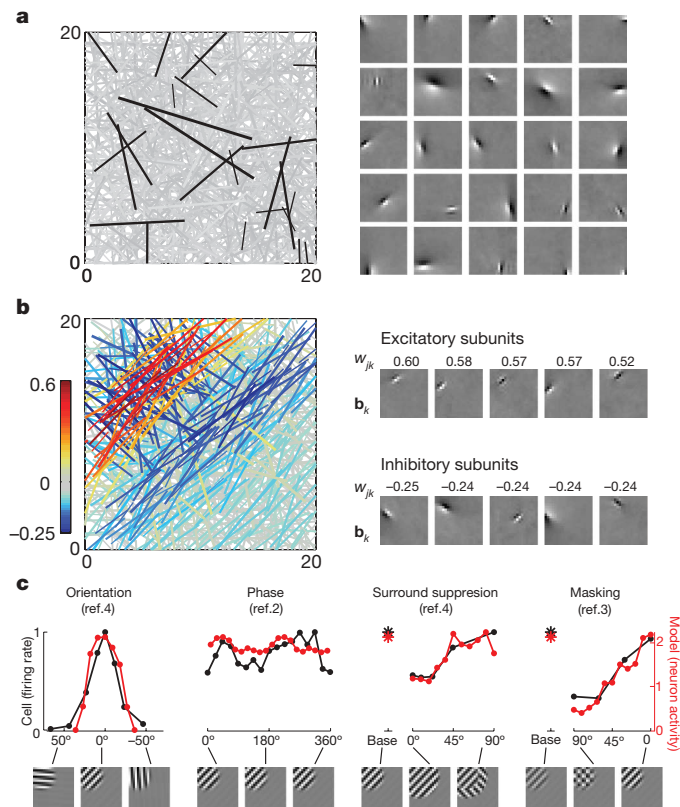
The vectors  $\mathbf{b}_k$  encode the directions of common expansion or contraction in the shape of the image distribution. Drawn as image patches, each is an oriented and localized edge-like feature. The full set tiles the spatial extent of the image patch (Fig. 3a) and spans the range of orientations and spatial frequencies of natural images (not shown). These oriented, band-pass image features are consistent with the optimal images for exciting simple cells in the primary visual cortex<sup>18,19</sup>. Similar representations have been derived previously using linear statistical models that maximize the efficiency of the image codes<sup>16,17</sup>. In the model proposed here, however, these features are not used explicitly to reconstruct the original image, but instead function to modify the encoded distributions (arrows in Fig. 2). Thus, whereas the traditional interpretation of early sensory codes is that they are adapted for faithful reconstruction of the stimulus, our results suggest an additional interpretation: they convey variations in image distributions and allow downstream visual areas to form more abstract representations.

The second set of parameters, the weights  $w_{jk}$ , describes the role of each neuron in shaping the encoded image distribution. A set of learned weights for a typical model neuron is shown in Fig. 3b. This neuron exerts the strongest effect on features in the top left of the image patch, increasing the variability (that is, activation) of those oriented at its 'preferred' orientation of 45°, decreasing the variability of those at the orthogonal orientation, as well as those at the preferred orientation but at an offset location. Rather than



responding to a few excitatory or suppressive image features, the neuron integrates a large number to describe a pattern of variability underlying a particular image distribution. Although the functional significance of these subunits is to modify the statistical structure of the encoded distribution, they also reflect stimulus features to which this model neuron is most sensitive. It should be noted that a model neuron is activated by all images from this distribution, rather than signalling the presence of one best stimulus. Conversely, stimuli that lie in parts of image space assigned low probability by the neuron inhibit its activity.

To compare the behaviour of the model neuron to that of cells in the visual cortex, we tested its response to stimuli used in classical



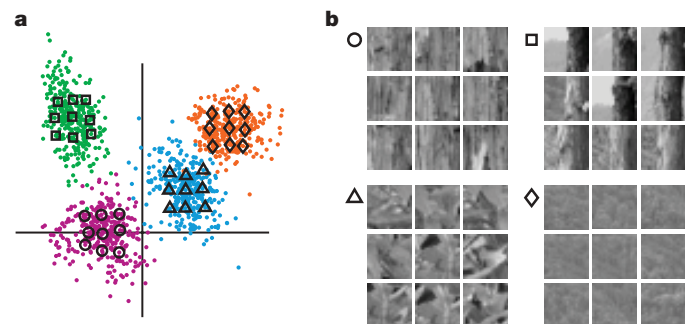
**Figure 3 | Model neurons exhibit properties of cortical visual neurons.**

**a**, When adapted to natural images, the vectors  $\mathbf{b}_k$  are oriented, localized in space, and span the spatial extent of the  $20 \times 20$ -pixel image patch. Each line reflects the orientation, spatial position within the image patch, and scale (length of line) of one of the image features. Twenty-five representative features (from a total of 1,000) are drawn in black, and shown in image form on the right. **b**, Weights of one typical model neuron to the features  $\mathbf{b}_k$ . As in **a**, each feature is represented by a line, and the colour of the line indicates the sign and magnitude of the weight to the feature (see colour bar). Positive weights indicate increased variability in the corresponding feature; negative weights indicate decreased variability; features to which the neuron is insensitive are shaded grey. Image features ( $\mathbf{b}_k$ ) corresponding to the five most positive and the five most negative weights are shown in the right panel; the corresponding weights are above each feature. These act as excitatory and inhibitory subunits for this neuron. **c**, When presented with sinusoidal gratings, this model neuron replicates common aspects of the neural response in complex cells in cortical area V1. It is highly tuned to the grating's orientation, but insensitive to its phase. Adding a grating into the surrounding region suppresses the response (third plot,  $0^\circ$ ) relative to baseline response to a single grating (asterisk), but this suppression is tuned to the orientation of the surround and is weakest when the surround is orthogonal to the preferred orientation ( $90^\circ$ ). Masking with a superimposed orthogonal grating suppresses the response (fourth plot,  $90^\circ$ ), but this suppression is also orientation-dependent. All model neuron responses are plotted on the same scale (red axis); cell firing rates in each plot were normalized to a maximum value of 1; preferred orientation was shifted to  $0^\circ$  for the model neuron and the cell in all plots.

physiological experiments (sinusoidal gratings). Model parameters were fixed after training on natural images, and neural response computed on a set of patterns centred in the visual area that evoked maximal response. This particular model neuron showed a variety of properties observed in complex cells in V1 and cells in V2, including phase invariance, orientation tuning and complex suppressive effects (Fig. 3c). A large subset of the population exhibits similar properties, whereas others encode more complex patterns that have been observed in higher visual areas V2 and V4 (a detailed analysis of the population and similarities to other experimental data are provided in the Supplementary Information). We emphasize that these results, as well as image features described in Fig. 3a, were obtained with no assumptions about the image structure encoded by visual neurons and without fitting a model to data from physiological experiments. Specifically, we did not restrict the encoded image features to be localized and oriented, nor did we prescribe in advance how the subunits are to be combined in the pattern represented by each neuron.

Finally, we looked at the way in which the model uses the population of neurons to represent images. If the model is able to generalize across the wide variability present in natural images, then image patches that are widely scattered in the original space of linear features should be tightly clustered in the space of the model's representation. This can be illustrated by projecting into two dimensions (as was done with image space in Fig. 1) the model representation of a collection of images (Fig. 4). As hypothesized, by encoding image distributions rather than the precise feature content of each image, model neurons are able to encode perceptually similar images with similar representations and to separate distinct image types.

One limitation of the statistical framework used here is that it does not furnish an explicit feed-forward algorithm for neural encoding. Nevertheless, it is possible to approximate inference in the model by a sequential feed-forward computation: a neuron integrates the squared responses of a large number of image features  $\mathbf{b}_k$  and correlates the pattern against its weights  $w_{jk}$  (see Supplementary Information for details). This computation can be viewed as a generalization of the standard model of complex cells, in which each complex cell takes as input the squared output of two simple cells<sup>9,10,20,21</sup>. In contrast, model neurons can receive many inputs, and the linear features themselves are learned. We find that the optimal number of input features varies greatly, and the features are integrated in a variety of ways. These predictions are consistent with recent analyses of functional subfields in V1 complex cells<sup>6,22</sup>. In addition, some model neurons integrate more complex spatial patterns (see Supplementary Information), which predicts a neural response to a richer variety of images than has been tested



**Figure 4 | Generalization across natural variability.** **a**, In contrast to linear projections (compare to Fig. 1b), a two-dimensional projection of the model's representation (the activity of 150 model neurons) reveals well-separated clusters. **b**, Each  $3 \times 3$ -image group corresponds to the array of symbols in **a**. Despite the variability in the appearance of edges and textures, the model's representation of natural images generalizes within each region while still distinguishing among them.

physiologically. Experiments that incorporate such stimuli will provide an important validation of the proposed model.

The nonlinear effects shown by neurons in the model (Fig. 3c) have been previously incorporated into models of complex cells<sup>5,8,20,21</sup>. Much of this work has focused on fitting mathematical models to neural data<sup>5,8,20,23</sup> and does not provide a functional explanation of the observed neural properties. Other models have been motivated by specific computational goals, such as statistical independence<sup>24,25</sup>, stability of representation over time<sup>26,27</sup>, or position or scale invariance<sup>28</sup>. However, these models do not explicitly address the problem of generalization, which here is performed by inferring the statistical distribution that is most likely to explain the input image. An important advantage of our approach is that, rather than assuming invariance (or sensitivity) to limited stimulus parameters such as position or orientation, the model learns a much more general set of features that are determined by the statistical structures in natural images. If higher-level visual neurons are generalizing according to these statistics, they should have invariance along specific stimulus dimensions, and their responses to natural images should reflect common statistical structure in local image regions. Thus, the model provides a quantitative way to explore neural responses to complex stimuli characterized by their statistical regularities.

## METHODS SUMMARY

The model describes individual image patches  $\mathbf{x}$  with multivariate Gaussian probability distributions:

$$P(\mathbf{x}|\mathbf{y}) = \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}) \quad (1)$$

with mean  $\boldsymbol{\mu} = 0$  and with covariance a function of the neural encoding vector  $\mathbf{C} = f(\mathbf{y})$ . The logarithm of the covariance matrix is given by the combination of outer products of feature vectors  $\mathbf{b}_k$ , weighted by neural activities  $y_j$  through weights  $w_{jk}$ :

$$\log \mathbf{C} = \sum_{jk} y_j w_{jk} \mathbf{b}_k \mathbf{b}_k^T \quad (2)$$

Because a different covariance can be inferred for each image, the distribution over the entire ensemble of images is highly non-Gaussian. (This model is a generalized version of the hierarchical model described previously<sup>19</sup>, which captured patterns among the variances, but not the correlations, of linear features.)

We trained the model on a large set of  $20 \times 20$  image patches, sampled randomly from greyscale photographs of outdoor scenes<sup>19</sup>. The number of neurons was set to 150 and the number of linear features  $\mathbf{b}_k$  to 1,000. The 'response' of model neurons was computed as the most probable neural representation given the input image by maximizing the posterior probability  $P(\mathbf{y}|\mathbf{x}, \{\mathbf{b}_k, w_{jk}\})$ . Model parameters were initialized to small random values and optimized by maximizing the likelihood of the data under the model  $P(\mathbf{x}|\{\mathbf{b}_k, w_{jk}\})$  using standard gradient ascent.

For the 'physiological' analysis of Fig. 3c, we first identified the location, orientation, and spatial extent and frequency of a windowed sinusoidal grating that best activated the model neuron (one that yielded the most positive value of  $y_j$ ). We then varied each tested parameter and computed the model's representation of the stimulus (the vector of responses of model neurons).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 4 May; accepted 26 September 2008.**

**Published online 19 November 2008.**

- Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol. (Lond.)* **160**, 106–154 (1962).
- Movshon, J. A., Thompson, I. D. & Tolhurst, D. J. Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *J. Physiol. (Lond.)* **283**, 53–77 (1978).
- Bonds, A. B. Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex. *Vis. Neurosci.* **2**, 41–55 (1989).

- Jones, H. E., Wang, W. & Sillito, A. M. Spatial organization and magnitude of orientation contrast interactions in primate V1. *J. Neurophysiol.* **88**, 2796–2808 (2002).
- Cavanaugh, J. R., Bair, W. & Movshon, J. A. Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *J. Neurophysiol.* **88**, 2530–2546 (2002).
- Chen, X., Han, F., Poo, M. & Dan, Y. Excitatory and suppressive receptive field subunits in awake monkey primary visual cortex (V1). *Proc. Natl Acad. Sci. USA* **104**, 19120–19125 (2007).
- Chichilnisky, E. J. A simple white noise analysis of neuronal light responses. *Network: Comp. Neural Syst.* **12**, 199–213 (2001).
- Carandini, M., Heeger, D. J. & Movshon, J. A. Linearity and normalization in simple cells of the macaque primary visual cortex. *J. Neurosci.* **17**, 8621–8644 (1997).
- Movshon, J. A., Thompson, I. D. & Tolhurst, D. J. Receptive field organization of complex cells in the cat's striate cortex. *J. Physiol. (Lond.)* **283**, 79–99 (1978).
- Adelson, E. H. & Bergen, J. R. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A* **2**, 284–299 (1985).
- Kobatake, E. & Tanaka, K. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J. Neurophysiol.* **71**, 856–867 (1994).
- Gallant, J. L., Connor, C. E., Rakshit, S., Lewis, J. W. & Van Essen, D. C. Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *J. Neurophysiol.* **76**, 2718–2739 (1996).
- Connor, C. E., Brincat, S. L. & Pasupathy, A. Transformation of shape information in the ventral pathway. *Curr. Opin. Neurobiol.* **17**, 140–147 (2007).
- Hegd , J. & Van Essen, D. C. Selectivity for complex shapes in primate visual area V2. *J. Neurosci.* **20**, RC61:1–6 (2000).
- Pasupathy, A. & Connor, C. E. Shape representation in area V4: position-specific tuning for boundary conformation. *J. Neurophysiol.* **86**, 2505–2519 (2001).
- Olshausen, B. A. & Field, D. J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607–609 (1996).
- Bell, A. J. & Sejnowski, T. J. The "independent components" of natural scenes are edge filters. *Vision Res.* **37**, 3327–3338 (1997).
- Jones, J. P. & Palmer, L. A. The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *J. Neurophysiol.* **58**, 1187–1211 (1987).
- van Hateren, J. H. & van der Schaaf, A. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. R. Soc. Lond. B* **265**, 359–366 (1998).
- Heeger, D. J. Normalization of cell responses in cat striate cortex. *Vis. Neurosci.* **9**, 181–197 (1992).
- Heeger, D. J., Simoncelli, E. P. & Movshon, J. A. Computational models of cortical visual processing. *Proc. Natl Acad. Sci. USA* **93**, 623–627 (1996).
- Rust, N. C., Schwartz, O., Movshon, J. A. & Simoncelli, E. P. Spatiotemporal elements of macaque V1 receptive fields. *Neuron* **46**, 945–956 (2005).
- Cadieu, C. et al. A model of V4 shape selectivity and invariance. *J. Neurophysiol.* **98**, 1733–1750 (2007).
- Schwartz, O. & Simoncelli, E. P. Natural signal statistics and sensory gain control. *Nature Neurosci.* **4**, 819–825 (2001).
- Hyv rinen, A. & Hoyer, P. A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images. *Vision Res.* **41**, 2413–2423 (2001).
- Berkes, P. & Wiskott, L. Slow feature analysis yields a rich repertoire of complex cell properties. *J. Vis.* **5**, 579–602 (2005).
- Hurri, J. & Hyv rinen, A. Simple-cell-like receptive fields maximize temporal coherence in natural video. *Neural Comput.* **15**, 663–691 (2003).
- Riesenhuber, M. & Poggio, T. Hierarchical models of object recognition in cortex. *Nature Neurosci.* **2**, 1019–1025 (1999).
- Karklin, Y. & Lewicki, M. S. A hierarchical Bayesian model for learning non-linear statistical regularities in non-stationary natural signals. *Neural Comput.* **17**, 397–423 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by the Department of Energy through the Computational Science Graduate Fellowship (to Y.K.), the National Science Foundation Grant under grant numbers 0413152 and 0705677 (to M.S.L.) and the Office of Naval Research under the Multidisciplinary University Research Initiative N000140710747.

**Author Contributions** Y.K. and M.S.L. developed the model, analysed the results and wrote the paper; Y.K. ran the simulations.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to Y.K. ([yan.karklin@nyu.edu](mailto:yan.karklin@nyu.edu)) or M.S.L. ([michael.lewicki@case.edu](mailto:michael.lewicki@case.edu)).



## METHODS

**Data.** We used 110 greyscale images of outdoor scenes as training data<sup>19</sup>. Pixel intensities were log-transformed (corresponding roughly to the transformation at the retinal cone cells<sup>30</sup>), and the images were low-pass filtered to remove corner frequency sampling artefacts. We randomly extracted overlapping  $20 \times 20$ -image patches from the entire data set. The mean luminance value was subtracted from each patch (which sped up model training but had no significant influence on the results). We ‘whitened’ all image patches to remove global correlations and to normalize the variance; this allowed the model to encode only the deviations of each image distribution from the global statistics (the canonical distribution). For visualization of image features, the results were projected back into the original image space. All stimuli in the physiological analysis of Fig. 3c were preprocessed in the same way as the natural images used in training.

**Model parameter estimation.** We estimated the optimal model parameters  $\theta = \{\mathbf{b}_k, w_{jk}\}$  by maximizing the likelihood of the data under the model

$$P(\mathbf{x}|\theta) = \int P(\mathbf{x}|\mathbf{y}, \theta) P(\mathbf{y}) d\mathbf{y} \quad (3)$$

The conditional distribution  $P(\mathbf{x}|\mathbf{y}, \theta)$  is a multivariate Gaussian that captures correlations in local image regions (equation (1)). Neural activities were assumed to be sparse<sup>31</sup> and independent, and were modelled with a Laplacian (symmetric exponential) distribution,  $P(\mathbf{y}) = \prod P(y_j) \propto \prod e^{-|y_j|}$ . The integral over all possible neural states in equation (3) is intractable and was replaced by a single evaluation at the maximum a posteriori value  $\hat{\mathbf{y}} = \arg \max_{\mathbf{y}} P(\mathbf{y}|\mathbf{x}, \theta)$ . Although this approximation ignores the volume around the maximum, it is one standard approach to tackling this problem.

We assumed the training patches were sampled independently and that the likelihood for the data ensemble was a product of terms for individual images (equation (3)). In practice, we maximized the log-likelihood using gradient ascent on batches of 100 image patches. Repeated training runs produced convergence to similar parameter values.

**Model responses to grating stimuli.** The orientation tuning of model neurons in Fig. 3c was measured using  $20 \times 20$  patches of sinusoidal gratings at different positions, orientations, spatial frequencies and phases. We first eliminated neurons that were ‘unresponsive’ to gratings, that is, those whose maximal response did not reach 2 standard deviations of the population response to gratings. This was necessary to discount small random activation of neurons tuned for other types of image structures. For each neuron we found the grating with the maximal response and measured modulation in response to varying orientation, phase, or the addition of masks in the receptive field or the surround. Because neural activity could be positive or negative, the full amplitude of modulation was considered as twice the maximum absolute value of the response.

A neuron was considered to be orientation-tuned if its response was modulated by more than 50% over the range of stimulus orientations, and to be phase invariant if the response varied less than 50% over phase-shifted gratings. Cross-orientation inhibition and surround suppression corresponded to greater than 25% decrease in neural response. Bandwidth of orientation tuning was computed as the width at  $1/\sqrt{2}$  of the full amplitude of the response modulation.

The projection of neural activity in Fig. 4 was computed using linear discriminant analysis, a technique that finds the linear projections that best separate different classes of data. Applied to the raw pixel data or to the outputs of linear features (data shown in Fig. 1), this method failed to separate the clusters.

30. van Hateren, J. H. Processing of natural time series of intensities by the visual system of the blowfly. *Vision Res.* **37**, 3407–3416 (1997).

31. Olshausen, B. A. & Field, D. J. Sparse coding of sensory inputs. *Curr. Opin. Neurobiol.* **14**, 481–487 (2004).

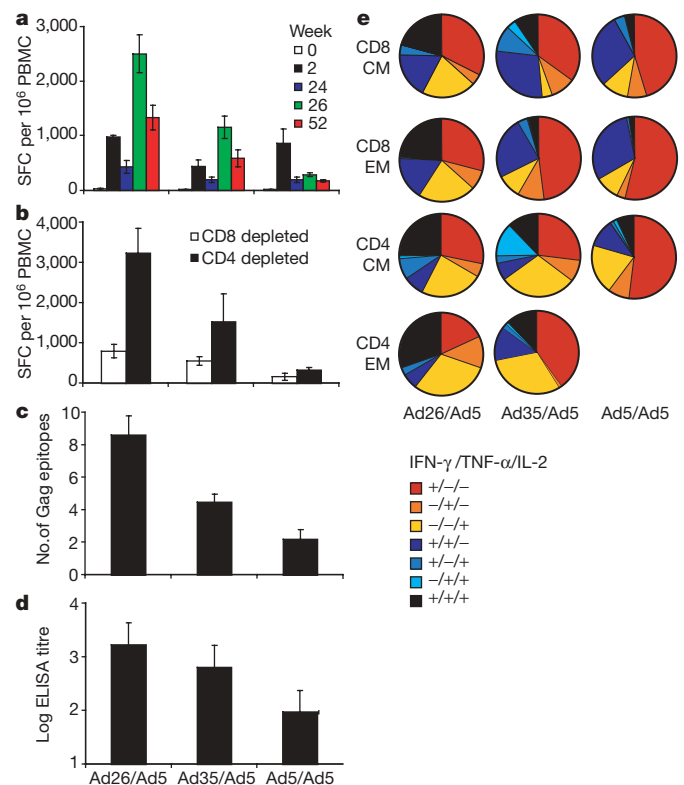
# Immune control of an SIV challenge by a T-cell-based vaccine in rhesus monkeys

Jinyan Liu<sup>1</sup>, Kara L. O'Brien<sup>1</sup>, Diana M. Lynch<sup>1</sup>, Nathaniel L. Simmons<sup>1</sup>, Annalena La Porte<sup>1</sup>, Ambryce M. Riggs<sup>1</sup>, Peter Abbink<sup>1</sup>, Rory T. Coffey<sup>1</sup>, Lauren E. Grandpre<sup>1</sup>, Michael S. Seaman<sup>1</sup>, Gary Landucci<sup>2</sup>, Donald N. Forthal<sup>2</sup>, David C. Montefiori<sup>3</sup>, Angela Carville<sup>4</sup>, Keith G. Mansfield<sup>4</sup>, Menzo J. Havenga<sup>5</sup>, Maria G. Pau<sup>6</sup>, Jaap Goudsmit<sup>6</sup> & Dan H. Barouch<sup>1</sup>

A recombinant adenovirus serotype 5 (rAd5) vector-based vaccine for HIV-1 has recently failed in a phase 2b efficacy study in humans<sup>1,2</sup>. Consistent with these results, preclinical studies have demonstrated that rAd5 vectors expressing simian immunodeficiency virus (SIV) Gag failed to reduce peak or setpoint viral loads after SIV challenge of rhesus monkeys (*Macaca mulatta*) that lacked the protective MHC class I allele *Mamu-A\*01* (ref. 3). Here we show that an improved T-cell-based vaccine regimen using two serologically distinct adenovirus vectors afforded substantially improved protective efficacy in this challenge model. In particular, a heterologous rAd26 prime/rAd5 boost vaccine regimen expressing SIV Gag elicited cellular immune responses with augmented magnitude, breadth and polyfunctionality as compared with the homologous rAd5 regimen. After SIV<sub>MAC251</sub> challenge, monkeys vaccinated with the rAd26/rAd5 regimen showed a 1.4log reduction of peak and a 2.4log reduction of setpoint viral loads as well as decreased AIDS-related mortality as compared with control animals. These data demonstrate that durable partial immune control of a pathogenic SIV challenge for more than 500 days can be achieved by a T-cell-based vaccine in *Mamu-A\*01*-negative rhesus monkeys in the absence of a homologous Env antigen. These findings have important implications for the development of next-generation T-cell-based vaccine candidates for HIV-1.

Recombinant Ad5 vector-based vaccines expressing SIV Gag have been shown to afford dramatic control of viral replication after simian-human immunodeficiency virus (SHIV)-89.6P challenge of rhesus monkeys<sup>4,5</sup>. However, rAd5-Gag vaccines have failed to reduce peak or setpoint viral loads after SIV<sub>MAC239</sub> challenge of rhesus monkeys<sup>3</sup>, highlighting important differences in the stringencies of these challenge models. Heterologous DNA prime/rAd5 boost vaccine regimens have also failed to date to reduce setpoint viral loads after SIV<sub>MAC</sub> challenge of rhesus monkeys that lacked the protective MHC class I allele *Mamu-A\*01* (refs 3 and 6). The inability of vector-based vaccines to afford durable control of setpoint viral loads after SIV<sub>MAC</sub> challenge of *Mamu-A\*01*-negative rhesus monkeys has led to substantial debate about the viability of the concept of developing T-cell-based vaccines for HIV-1.

Pre-existing Ad5-specific neutralizing antibodies have been reported to reduce the immunogenicity of rAd5 vector-based vaccines in clinical trials<sup>7,8</sup> and may also compromise their safety<sup>1</sup>. Rare serotype rAd vectors—such as rAd26 and rAd35 vectors<sup>9–12</sup>—have been developed as potential alternatives. Serologically distinct rAd vectors also allow the development of more potent heterologous rAd



**Figure 1 | Immunogenicity of heterologous rAd prime-boost vaccine regimens.** Rhesus monkeys were primed at week 0 and boosted at week 24 with rAd26/rAd5, rAd35/rAd5 or rAd5/rAd5 regimens expressing SIV Gag. **a**, Gag-specific IFN- $\gamma$  ELISPOT assays were performed at weeks 0, 2, 24, 26 and 52 after immune priming. **b**, CD4<sup>+</sup> (white bars) and CD8<sup>+</sup> (black bars) T lymphocyte responses were evaluated at week 28 by CD8-depleted and CD4-depleted ELISPOT assays, respectively. SFC, spot-forming cells. **c**, The breadth of responses was determined by Gag epitope mapping at week 28. **d**, Gag-specific antibody responses were determined by ELISA at week 28. Mean responses with standard errors are shown (**a–d**). **e**, Functionality of Gag-specific CD8<sup>+</sup> and CD4<sup>+</sup> central memory (CM; CD28<sup>+</sup>CD95<sup>+</sup>) and effector memory (EM; CD28<sup>−</sup>CD95<sup>+</sup>) T lymphocyte responses were assessed by 8-colour ICS assays. Proportions of IFN- $\gamma$ , TNF- $\alpha$  and IL-2 responses are depicted individually and in all possible combinations for each cellular subpopulation. CD4<sup>+</sup> EM responses after rAd5/rAd5 immunization were below the detection limit of the assay.

<sup>1</sup>Division of Viral Pathogenesis, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, USA. <sup>2</sup>University of California, Irvine School of Medicine, Irvine, California 92697, USA. <sup>3</sup>Duke University Medical Center, Durham, North Carolina 27710, USA. <sup>4</sup>New England Primate Research Center, Southborough, Massachusetts 01772, USA. <sup>5</sup>TNO Biosciences, 2301 CE, Leiden, The Netherlands. <sup>6</sup>Cruell Holland BV, 2301 CA, Leiden, The Netherlands.



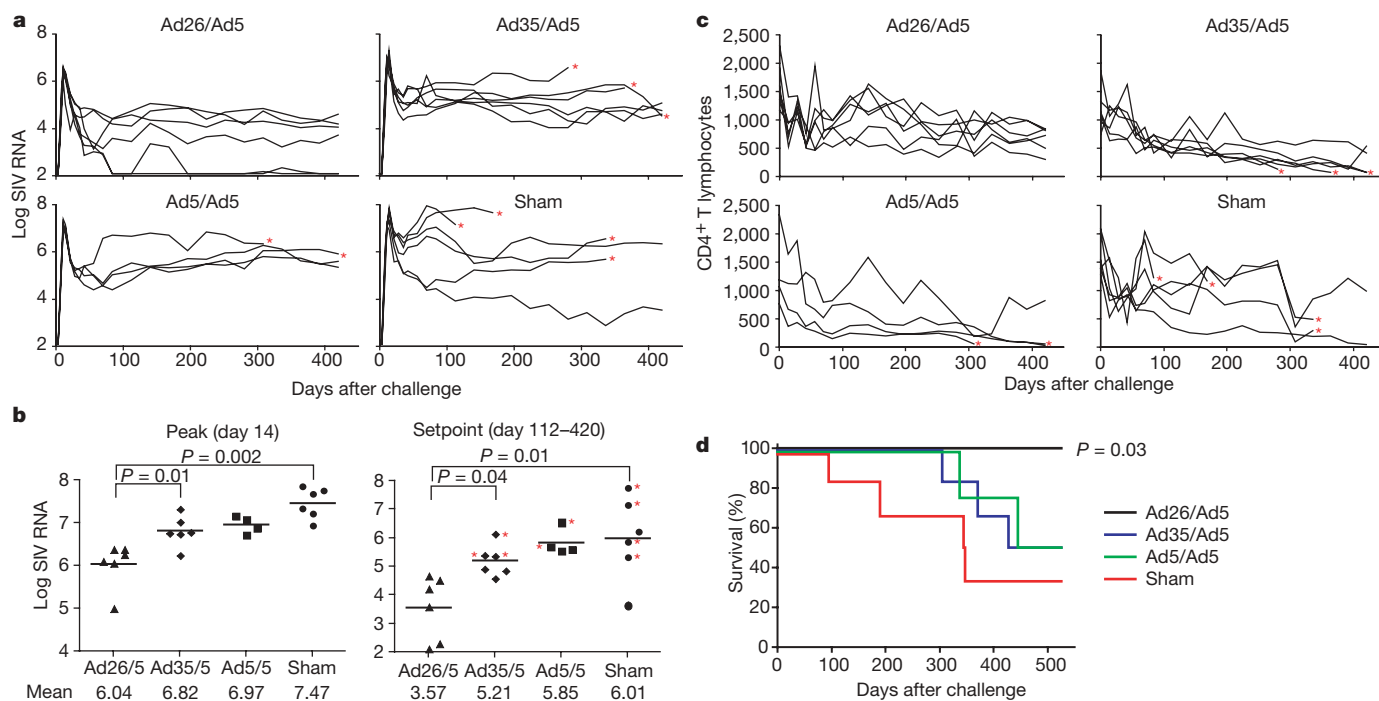
prime-boost vaccine regimens<sup>11</sup>. To investigate the immunogenicity and protective efficacy of such regimens, we immunized 22 Indian-origin rhesus monkeys (*Macaca mulatta*) that lacked the protective MHC class I alleles *Mamu-A\*01* (refs 13–15) and *Mamu-B\*17* (ref. 16) with the following heterologous or homologous rAd prime-boost regimens: (1) rAd26-Gag prime/rAd5-Gag boost ( $n = 6$ ); (2) rAd35-Gag prime/rAd5-Gag boost ( $n = 6$ ); (3) rAd5-Gag prime/rAd5-Gag boost ( $n = 4$ ); and (4) sham controls ( $n = 6$ ). One monkey each in groups 1, 3 and 4 expressed the protective *Mamu-B\*08* allele. Monkeys were primed at week 0 and boosted at week 24 with  $10^{11}$  viral particles of each vector expressing SIV<sub>MAC239</sub> Gag. At week 52, all animals received a high-dose intravenous (i.v.) challenge with 100 infectious doses of SIV<sub>MAC251</sub> (ref. 6).

We monitored vaccine-elicited SIV Gag-specific cellular (Fig. 1a–c) and humoral (Fig. 1d) immune responses in these animals before challenge. After the priming immunization, IFN- $\gamma$  enzyme-linked immunospot (ELISPOT) responses to pooled SIV Gag peptides were observed in all vaccinees. Monkeys primed with rAd26-Gag or rAd35-Gag were efficiently boosted by the heterologous rAd5-Gag vector to peak responses of 2,513 and 1,163 spot-forming cells per  $10^6$  peripheral blood mononuclear cells, respectively, 2 weeks after the boost immunization (Fig. 1a; green bars). In contrast, monkeys primed with rAd5-Gag were only marginally boosted by a second injection of rAd5-Gag as a result of anti-vector immunity generated by the priming immunization<sup>11,17</sup>. Cell-depleted ELISPOT assays demonstrated that these responses were primarily CD8<sup>+</sup> T lymphocyte responses, although lower levels of CD4<sup>+</sup> T lymphocyte responses were also clearly observed (Fig. 1b). Epitope mapping was then performed by assessing ELISPOT responses against all 125 individual 15-amino-acid SIV Gag peptides after the boost immunization. The rAd26/rAd5 regimen elicited a mean of 8.6 detectable Gag epitopes per animal, whereas the rAd35/rAd5 regimen elicited a mean of 4.5 epitopes per animal and the rAd5/rAd5 regimen induced a mean of only 2.2 epitopes per animal (Fig. 1c). These data demonstrate that the heterologous rAd26/rAd5 regimen induced an

8.7-fold greater magnitude and a 3.9-fold increased breadth of Gag-specific cellular immune responses as compared with the homologous rAd5/rAd5 regimen.

We next assessed the functionality of the vaccine-elicited, Gag-specific T lymphocyte responses by multiparameter flow cytometry<sup>18,19</sup>. Intracellular cytokine staining (ICS) assays were performed to assess IFN- $\gamma$ , TNF- $\alpha$  and IL-2 secretion in CD8<sup>+</sup> and CD4<sup>+</sup> central memory (CD28<sup>+</sup>CD95<sup>+</sup>) and effector memory (CD28<sup>−</sup>CD95<sup>+</sup>) T lymphocyte subpopulations<sup>20,21</sup>. Consistent with our previous findings<sup>11</sup>, we observed larger proportions of polyfunctional IFN- $\gamma$ /TNF- $\alpha$ /IL-2-positive (black) and IL-2-positive (yellow) CD8<sup>+</sup> and CD4<sup>+</sup> T lymphocyte responses in the animals that received the heterologous rAd26/rAd5 regimen as compared with the homologous rAd5/rAd5 regimen (Fig. 1e). In contrast, the rAd5/rAd5 regimen elicited primarily IFN- $\gamma$ -positive (red) and IFN- $\gamma$ /TNF- $\alpha$ -positive (blue) responses.

Six months after the boost immunization, all animals were challenged intravenously with SIV<sub>MAC251</sub>. Protective efficacy was evaluated by monitoring plasma SIV RNA levels (Fig. 2a, b), CD4<sup>+</sup> T lymphocyte counts (Fig. 2c), and clinical disease progression and mortality (Fig. 2d) after challenge. Monkeys that received the rAd26/rAd5 regimen showed mean peak viral loads that were 1.43 logs lower than the mean peak viral loads in the control animals (Fig. 2a, b;  $P = 0.002$ , two-sided Wilcoxon's rank-sum test with multiple comparison adjustments). More importantly, the rAd26/rAd5 vaccinated animals also demonstrated durable partial control of setpoint viral loads, as defined as the mean SIV RNA levels from day 112 to day 420 after challenge. In particular, the rAd26/rAd5-vaccinated monkeys maintained a 2.44 log reduction of setpoint viral loads as compared with the control animals (Fig. 2a, b;  $P = 0.01$ ). If the three *Mamu-B\*08*-positive animals are excluded from the analysis, then the rAd26/rAd5 regimen afforded a 2.66 log reduction of setpoint viral loads as compared with the control animals ( $P = 0.008$ ; data not shown). As expected, the homologous rAd5/rAd5 regimen afforded no detectable control of setpoint viral loads, consistent with



**Figure 2 | Protective efficacy of heterologous rAd prime-boost vaccine regimens.** Monkeys were challenged intravenously with SIV<sub>MAC251</sub>, and protective efficacy was monitored by SIV RNA levels (**a**, **b**), CD4<sup>+</sup> T lymphocyte counts (**c**), and clinical disease progression and mortality (**d**) after challenge. Viral loads are depicted longitudinally for each group

(**a**), and peak (day 14) and setpoint (day 112–420) viral loads are summarized for each group (**b**). Asterisks indicate mortality. Comparisons among groups were performed by two-sided Wilcoxon's rank-sum (**a**, **b**) and Fisher's exact (**d**) tests.

the previously reported failure of this regimen in this stringent SIV challenge model<sup>3</sup>.

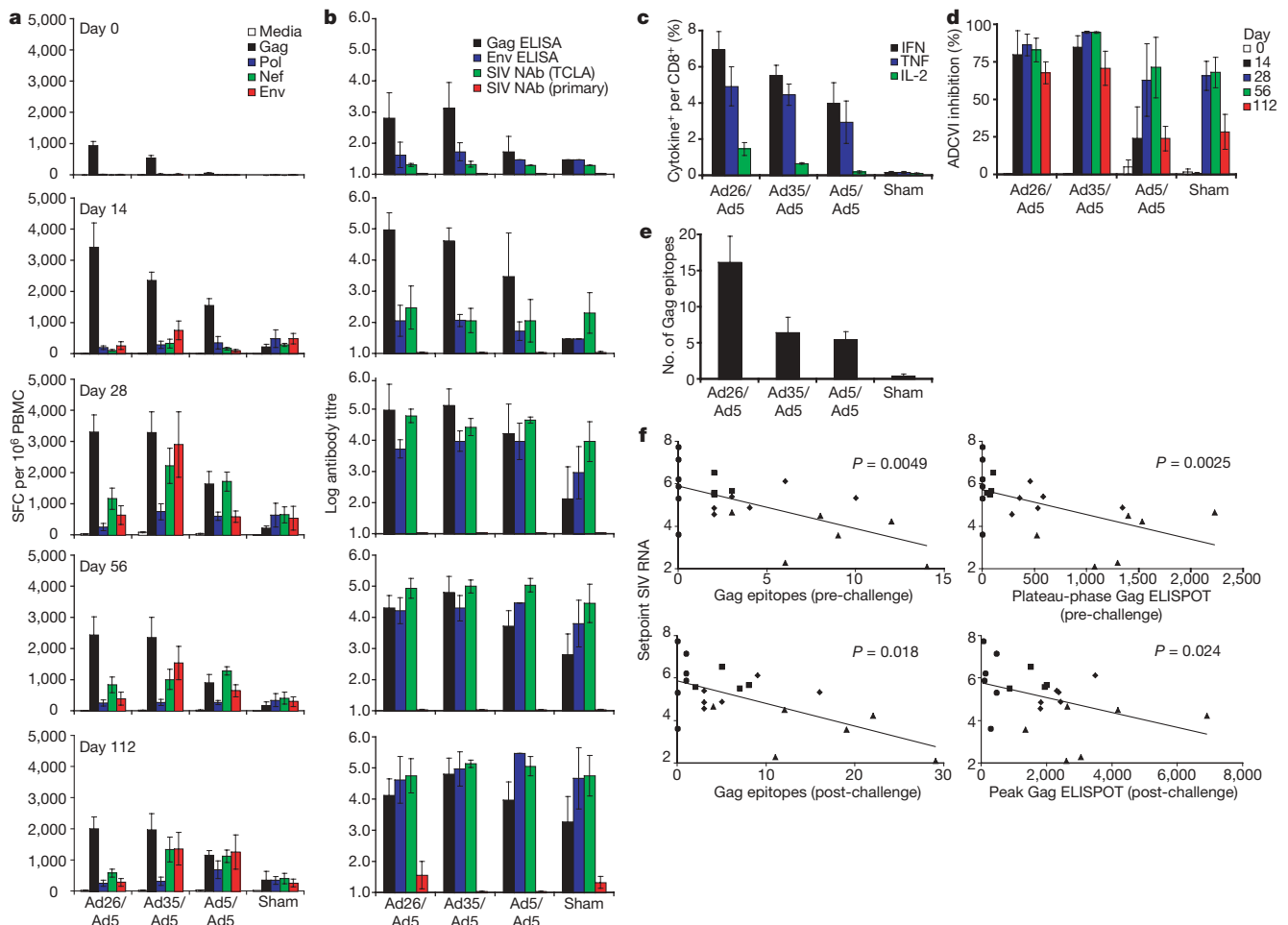
Monkeys that received the rAd26/rAd5 regimen also demonstrated slower declines of CD4<sup>+</sup> T lymphocyte counts as compared with the other groups (Fig. 2c). Moreover, the rAd26/rAd5 vaccine afforded a significant reduction of AIDS-related mortality as compared with the controls at day 500 after challenge (Fig. 2d;  $P = 0.03$ , Fisher's exact test), whereas the rAd35/rAd5 and rAd5/rAd5 vaccines provided a trend towards a survival advantage. Necropsies showed that the causes of death were AIDS-defining illnesses, including *Pneumocystis pneumonia*, *Cryptosporidium enteritis*, cytomegalovirus pneumonia, SIV encephalitis and lymphoma.

We next evaluated the evolution of SIV-specific cellular (Fig. 3a, c, e) and humoral (Fig. 3b, d) immune responses in these animals after challenge. The rAd26/rAd5 vaccinees exhibited tenfold greater anamnestic Gag-specific ELISPOT responses as compared with the controls during both acute and chronic infection (Fig. 3a; black bars), suggesting the critical importance of these responses for immune control. In contrast, Pol-, Nef- and Env-specific cellular immune responses were either comparable or lower in magnitude in the rAd26/rAd5 vaccinees as compared with the other groups (Fig. 3a). The rAd26/rAd5 vaccinees thus showed less extensive diversification of cellular immune responses against these other SIV antigens,

presumably reflecting the reduced viral replication in these animals. Phenotypic analysis of the Gag-specific CD8<sup>+</sup> T lymphocyte responses in the rAd26/rAd5 vaccinees demonstrated slightly higher expression of IFN- $\gamma$  and TNF- $\alpha$  but 6.7-fold greater expression of IL-2 as compared with the rAd5/rAd5 vaccinees after challenge (Fig. 3c; green bars). These data suggest that not only quantitative but also qualitative differences among the vaccine regimens may have contributed to protective efficacy.

We assessed the breadth of Gag-specific cellular immune responses on day 28 after challenge by epitope mapping using 15-amino-acid SIV Gag peptides (Fig. 3e). The monkeys that received the rAd26/rAd5 regimen developed a mean of 16.1 (range 4–29) Gag epitope-specific responses after challenge. In contrast, the control animals developed <1 detectable Gag epitope-specific response after challenge. Notably, both the breadth and the magnitude of Gag-specific cellular immune responses elicited by vaccination ( $P = 0.0049$  and  $P = 0.0025$ , respectively, Spearman's rank-correlation tests) and recalled after challenge ( $P = 0.018$  and  $P = 0.024$ , respectively) correlated with control of setpoint viral loads in these animals (Fig. 3f).

Humoral immune responses were evaluated by Gag- and Env-specific ELISAs as well as by luciferase-based pseudovirus neutralizing antibody assays using both laboratory-adapted and primary isolate SIV<sub>MAC251</sub> viruses<sup>22</sup> (Fig. 3b). The vaccinees developed more



**Figure 3 | Cellular and humoral immune responses after challenge.**

Cellular (a, c, e) and humoral (b, d) immune responses were assessed after challenge. IFN- $\gamma$  ELISPOT assays (a) and ELISA and neutralizing antibody assays (b) were performed on days 0, 14, 28, 56 and 112 after challenge. NAB, neutralizing antibody; primary, primary isolate virus; TCLA, T-cell laboratory-adapted virus. Eight-colour ICS assays on day 28 (c), ADCVI assays at 1:100 serum dilution at the time points indicated (d), and Gag epitope mapping studies on day 28 (e) were also performed. Mean responses

with standard errors are shown (a–e). f, Correlations between the breadth (left panels) or magnitude (right panels) of pre-challenge (top panels) or post-challenge (bottom panels) Gag-specific cellular immune responses and setpoint viral loads were evaluated by two-sided Spearman's rank correlation tests. Monkeys immunized with rAd26/rAd5 (triangles), rAd35/rAd5 (diamonds), rAd5/rAd5 (squares) and sham (circles) vaccine regimens are depicted.



rapid kinetics of Gag-specific ELISA responses as compared with the controls after challenge, but no clear differences of Env-specific ELISA or neutralizing antibody responses were observed among groups. Compared with controls, the vaccinees also developed more rapid kinetics of antibody-dependent cell-mediated virus inhibition (ADCVI)<sup>23</sup> on day 14 after challenge (Fig. 3d;  $P = 0.01$ ), suggesting the potential functional relevance of nonclassical antibody responses in early SIV infection.

We next monitored CD4<sup>+</sup> T lymphocyte dynamics in these animals after challenge. The monkeys that received the rAd26/rAd5 regimen had a relative preservation of central memory CD4<sup>+</sup> T lymphocytes<sup>6</sup> (Fig. 4a; black bars;  $P = 0.02$ , Wilcoxon's rank-sum test) as well as a more marked preservation of CCR5<sup>+</sup> central memory CD4<sup>+</sup> T lymphocytes<sup>24,25</sup> (Fig. 4b; black bars;  $P = 0.002$ ) as compared with the controls on day 14 after challenge. The rAd26/rAd5 vaccinees also showed reduced Ki67<sup>+</sup> proliferation of CCR5<sup>+</sup> central memory (Fig. 4c, d;  $P = 0.004$ ) and effector memory (Fig. 4e, f;  $P = 0.004$ ) CD4<sup>+</sup> T lymphocytes as compared with the controls after challenge. In addition, on day 21 after challenge, monkeys that received the rAd26/rAd5 regimen maintained markedly higher levels of gastrointestinal CD4<sup>+</sup> T lymphocytes in duodenal biopsies as compared with the controls, which showed extensive depletion of this lymphocyte

population as expected<sup>26–28</sup> (Fig. 4g;  $P = 0.02$ , Wilcoxon's rank-sum test).

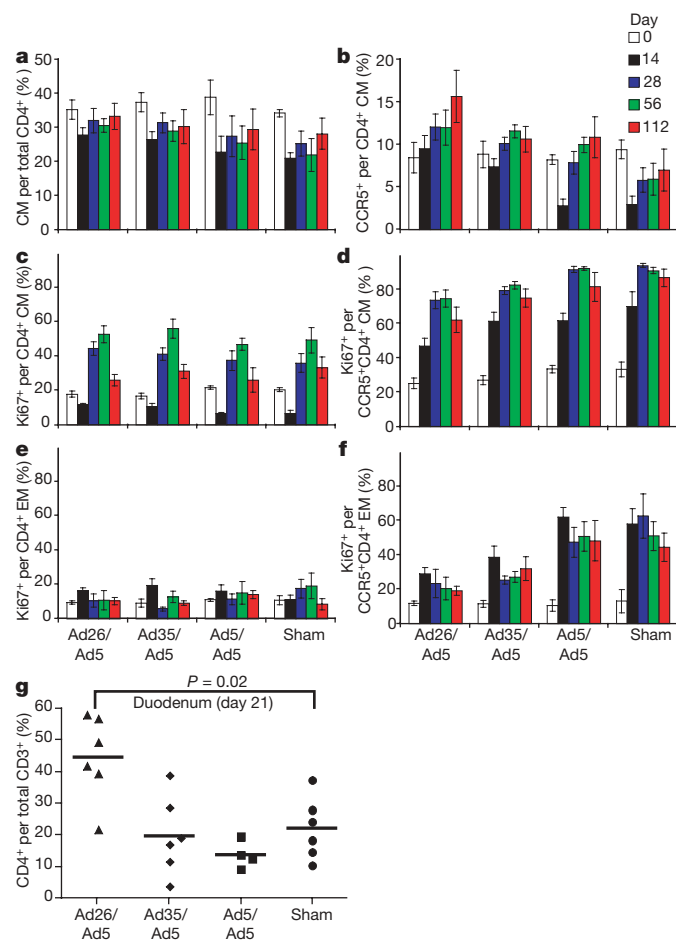
Our data demonstrate that the heterologous rAd26/rAd5 regimen elicited improved magnitude, breadth and functionality of Gag-specific cellular immune responses as compared with the homologous rAd5/rAd5 regimen and afforded durable partial immune control of a homologous SIV<sub>MAC251</sub> challenge in rhesus monkeys. To the best of our knowledge, vector-based vaccines have not previously been reported to reduce setpoint viral loads in the stringent system of SIV<sub>MAC</sub> challenge of *Mamu-A\*01*-negative rhesus monkeys<sup>3,6</sup>, although a previous study has shown partial control of setpoint viral loads using a DNA/rAd5 regimen in the less stringent system of *Mamu-A\*01*-positive rhesus monkeys<sup>29</sup>. In the present study, the breadth and magnitude of vaccine-elicited, Gag-specific cellular immune responses before challenge correlated with control of setpoint viral loads after challenge (Fig. 3f), suggesting the critical importance of Gag-specific T lymphocyte responses in controlling viral replication. Protective efficacy was also associated with improved functionality of Gag-specific T lymphocyte responses (Figs 1e and 3c), reduced diversification of responses against other SIV antigens after challenge (Fig. 3a), and preservation of CCR5<sup>+</sup> central memory CD4<sup>+</sup> T lymphocytes and gastrointestinal CD4<sup>+</sup> T lymphocytes after challenge (Fig. 4b, g), although some of these parameters may reflect the result rather than the cause of the reduced viral replication.

It is important to highlight the fact that the vaccines used in the present study expressed only a single SIV Gag antigen and did not include a homologous Env immunogen. The observed protective efficacy was therefore presumably mediated by vaccine-elicited Gag-specific cellular immune responses, because it is unlikely that Gag-specific antibodies afforded substantial protection. Consistent with these data, a previous study reported that the breadth of Gag-specific cellular immune responses correlated with control of viral loads in chronically HIV-1-infected humans<sup>30</sup>. It remains possible, however, that other SIV antigens may also contribute to cellular immune protection if included in a multivalent vaccine. Future studies should address this possibility and evaluate the protective efficacy of optimal vaccine regimens against highly heterologous SIV challenges. The potential role of early ADCVI responses should also be explored, because ADCVI activity emerged more rapidly in the vaccinees as compared with the controls after challenge (Fig. 3d). These responses probably reflect immune preservation in the vaccinees rather than a direct result of vaccination, because ADCVI activity has been reported to be mediated by Env-specific antibodies<sup>23</sup>.

Despite current controversies about the use of rAd vector-based vaccines for HIV-1, our data have important implications for the development of next generation T-cell-based vaccines by the proof-of-concept demonstration that durable partial immune control of a pathogenic SIV challenge can be achieved in *Mamu-A\*01*-negative rhesus monkeys. It remains unclear, however, whether the observed protection reflected the use of rAd26 vectors, the heterologous rAd prime-boost regimen, or both. Further studies will therefore be required to evaluate these possibilities and to determine the use of regimens consisting of two rare serotype rAd vectors. In conclusion, our findings suggest that T-cell-based vaccine regimens that elicit augmented magnitude, breadth and polyfunctionality of cellular immune responses as compared with the homologous rAd5 regimen may afford superior protective efficacy against HIV-1 and other pathogens.

## METHODS SUMMARY

**Animals, immunizations and challenge.** Outbred rhesus monkeys that did not express the MHC class I alleles *Mamu-A\*01* and *Mamu-B\*17* were housed at New England Primate Research Center (NEPRC), Southborough, Massachusetts, USA. Immunizations involved 10<sup>11</sup> viral particles of replication-incompetent, E1/E3-deleted rAd5, rAd35 or rAd26 vectors<sup>10–12</sup> expressing SIV<sub>MAC239</sub> Gag and were delivered as 1-ml injections intramuscularly in both quadriceps muscles at weeks 0 and 24. Sham controls received 10<sup>11</sup> viral particles



**Figure 4 | CD4<sup>+</sup> T lymphocyte dynamics after challenge.** Preservation of central memory (CM; CD28<sup>+</sup>CD95<sup>+</sup>) CD4<sup>+</sup> T lymphocytes (a) and CCR5<sup>+</sup> central memory CD4<sup>+</sup> T lymphocytes (b) was assessed on days 0, 14, 28, 56 and 112 after challenge. Ki67 staining of central memory CD4<sup>+</sup> (c), CCR5<sup>+</sup> central memory CD4<sup>+</sup> (d), effector memory (EM; CD28<sup>−</sup>CD95<sup>+</sup>) CD4<sup>+</sup> (e), and CCR5<sup>+</sup> effector memory CD4<sup>+</sup> (f) T lymphocytes was also determined. Mean responses with standard errors are shown (a–f). g, Preservation of gastrointestinal CD4<sup>+</sup> T lymphocytes was assessed in duodenal biopsies on day 21 after challenge. Comparisons were performed by two-sided Wilcoxon's rank-sum tests.

of empty vectors. At week 52, all animals received an i.v. inoculation of 100 infectious doses of SIV<sub>MAC251</sub> as described<sup>6</sup>. SIV RNA levels were assessed after challenge on days 0, 3, 7, 10, 14 and 21, then weekly until day 42, biweekly until day 84, and monthly thereafter (Siemens Diagnostics). All animal studies were approved by our Institutional Animal Care and Use Committees (IACUC).

**Cellular immune assays.** SIV-specific cellular immune responses were assessed by IFN- $\gamma$  ELISPOT assays and multiparameter ICS assays essentially as described<sup>11,20,21</sup>. ELISPOT assays used pooled or individual SIV Gag peptides. Eight-colour ICS assays used pooled SIV Gag peptides and the following monoclonal antibodies (BD Pharmingen): anti-CD3-Alexa700 (SP34), anti-CD4-AmCyan (L200), anti-CD8-APC-Cy7 (SK1), anti-CD28-PerCP-Cy5.5 (L293), anti-CD95-PE (DX2), anti-IFN- $\gamma$ -PE-Cy7 (B27), anti-IL-2-APC (MQ1-17H12) and anti-TNF- $\alpha$ -FITC (Mab11). Assessment of T lymphocyte dynamics used the following monoclonal antibodies: anti-CD3-Alexa700 (SP34), anti-CD4-AmCyan (L200), anti-CD8-APC-Cy7 (SK1), anti-CD28-PerCP-Cy5.5 (L293), anti-CD95-APC (DX2), anti-CCR5-PE (3A9), anti-HLA-DR-PE-Cy7 (L243) and anti-Ki67-FITC (B56). Gastrointestinal CD4<sup>+</sup> T lymphocytes were evaluated after collagenase digestion of duodenal biopsies and percoll gradient purification.

**Humoral immune assays.** SIV-specific humoral immune responses were evaluated by SIV Gag- and SIV Env gp130-specific ELISAs, luciferase-based pseudovirus neutralization assays (using T-cell laboratory adapted and primary isolate viruses)<sup>22</sup>, and ADCVI assays<sup>23</sup> essentially as described.

Received 18 August; accepted 29 September 2008.

Published online 9 November 2008.

- Fauci, A. S. The release of new data from the HVTN 502 (STEP) HIV vaccine study. National Institutes of Health (NIH) News, November 7 (2007).
- Fauci, A. S. 25 years of HIV. *Nature* **453**, 289–290 (2008).
- Casimiro, D. R. *et al.* Attenuation of simian immunodeficiency virus SIVmac239 infection by prophylactic immunization with DNA and recombinant adenoviral vaccine vectors expressing Gag. *J. Virol.* **79**, 15547–15555 (2005).
- Shiver, J. W. *et al.* Replication-incompetent adenoviral vaccine vector elicits effective anti-immunodeficiency-virus immunity. *Nature* **415**, 331–335 (2002).
- Shiver, J. W. & Emini, E. A. Recent advances in the development of HIV-1 vaccines using replication-incompetent adenovirus vectors. *Annu. Rev. Med.* **55**, 355–372 (2004).
- Letvin, N. L. *et al.* Preserved CD4<sup>+</sup> central memory T cells and survival in vaccinated SIV-challenged monkeys. *Science* **312**, 1530–1533 (2006).
- Catanzaro, A. T. *et al.* Phase 1 safety and immunogenicity evaluation of a multiclade HIV-1 candidate vaccine delivered by a replication-defective recombinant adenovirus vector. *J. Infect. Dis.* **194**, 1638–1649 (2006).
- Priddy, F. H. *et al.* Safety and immunogenicity of a replication-incompetent adenovirus type 5 HIV-1 clade B gag/pol/nef vaccine in healthy adults. *Clin. Infect. Dis.* **46**, 1769–1781 (2008).
- Barouch, D. H. *et al.* Immunogenicity of recombinant adenovirus serotype 35 vaccine in the presence of pre-existing anti-Ad5 immunity. *J. Immunol.* **172**, 6290–6297 (2004).
- Abbink, P. *et al.* Comparative seroprevalence and immunogenicity of six rare serotype recombinant adenovirus vaccine vectors from subgroups B and D. *J. Virol.* **81**, 4654–4663 (2007).
- Liu, J. *et al.* Magnitude and phenotype of cellular immune responses elicited by recombinant adenovirus vectors and heterologous prime-boost regimens in rhesus monkeys. *J. Virol.* **82**, 4844–4852 (2008).
- Vogels, R. *et al.* Replication-deficient human adenovirus type 35 vectors for gene transfer and vaccination: efficient human cell infection and bypass of preexisting adenovirus immunity. *J. Virol.* **77**, 8263–8271 (2003).
- Mothe, B. R. *et al.* Expression of the major histocompatibility complex class I molecule Mamu-A\*01 is associated with control of simian immunodeficiency virus SIV<sub>mac239</sub> replication. *J. Virol.* **77**, 2736–2740 (2003).
- Pal, R. *et al.* ALVAC-SIV-gag-pol-env-based vaccination and macaque major histocompatibility complex class I (A\*01) delay simian immunodeficiency virus SIV<sub>mac</sub>-induced immunodeficiency. *J. Virol.* **76**, 292–302 (2002).
- Zhang, Z. Q. *et al.* Mamu-A\*01 allele-mediated attenuation of disease progression in simian-human immunodeficiency virus infection. *J. Virol.* **76**, 12845–12854 (2002).
- Yant, L. J. *et al.* The high-frequency major histocompatibility complex class I allele Mamu-B\*17 is associated with control of simian immunodeficiency virus SIVmac239 replication. *J. Virol.* **80**, 5074–5077 (2006).
- Santra, S. *et al.* Replication-defective adenovirus serotype 5 vectors elicit durable cellular and humoral immune responses in nonhuman primates. *J. Virol.* **79**, 6516–6522 (2005).
- Betts, M. R. *et al.* HIV nonprogressors preferentially maintain highly functional HIV-specific CD8<sup>+</sup> T cells. *Blood* **107**, 4781–4789 (2006).
- Darrah, P. A. *et al.* Multifunctional T<sub>H</sub>1 cells define a correlate of vaccine-mediated protection against *Leishmania major*. *Nature Med.* **13**, 843–850 (2007).
- Okoye, A. *et al.* Progressive CD4<sup>+</sup> central memory T cell decline results in CD4<sup>+</sup> effector memory insufficiency and overt disease in chronic SIV infection. *J. Exp. Med.* **204**, 2171–2185 (2007).
- Pitcher, C. J. *et al.* Development and homeostasis of T cell memory in rhesus macaque. *J. Immunol.* **168**, 29–43 (2002).
- Montefiori, D. *Evaluating Neutralizing Antibodies Against HIV, SIV and SHIV in Luciferase Reporter Gene Assays* (eds Coligan, J. E *et al.*) (Current Protocols in Immunology, John Wiley & Sons, 2004).
- Forthal, D. N., Landucci, G. & Daar, E. S. Antibody from patients with acute human immunodeficiency virus (HIV) infection inhibits primary strains of HIV type 1 in the presence of natural-killer effector cells. *J. Virol.* **75**, 6953–6961 (2001).
- Veazey, R. S. *et al.* Dynamics of CCR5 expression by CD4<sup>+</sup> T cells in lymphoid tissues during simian immunodeficiency virus infection. *J. Virol.* **74**, 11001–11007 (2000).
- Veazey, R. S. *et al.* Increased loss of CCR5<sup>+</sup>CD45RA<sup>+</sup>CD4<sup>+</sup> T cells in CD8<sup>+</sup> lymphocyte-depleted Simian immunodeficiency virus-infected rhesus monkeys. *J. Virol.* **82**, 5618–5630 (2008).
- Veazey, R. S. *et al.* Gastrointestinal tract as a major site of CD4<sup>+</sup> T cell depletion and viral replication in SIV infection. *Science* **280**, 427–431 (1998).
- Li, Q. *et al.* Peak SIV replication in resting memory CD4<sup>+</sup> T cells depletes gut lamina propria CD4<sup>+</sup> T cells. *Nature* **434**, 1148–1152 (2005).
- Mattapallil, J. J. *et al.* Massive infection and loss of memory CD4<sup>+</sup> T cells in multiple tissues during acute SIV infection. *Nature* **434**, 1093–1097 (2005).
- Wilson, N. A. *et al.* Vaccine-induced cellular immune responses reduce plasma viral concentrations after repeated low-dose challenge with pathogenic simian immunodeficiency virus SIVmac239. *J. Virol.* **80**, 5875–5885 (2006).
- Kiepiela, P. *et al.* CD8<sup>+</sup> T-cell responses to different HIV proteins have discordant associations with viral load. *Nature Med.* **13**, 46–53 (2007).

**Acknowledgements** We thank R. Vogels, J. Custers, L. Holterman, A. Lemckert, F. Stephens, R. Dolin, N. Letvin, J. Schmitz, M. Lifton, K. Furr, L. Picker and M. Pensiero for generous advice, assistance and reagents. The SIV peptide pools were obtained from the National Institutes of Health (NIH) AIDS Research and Reference Reagent Program. We acknowledge support from NIH grants AI066305 (D.H.B.), AI066924 (D.H.B.), AI078526 (D.H.B.), AI030034 (D.C.M.) and RR000168.

**Author Contributions** J.L., K.L.O'B., D.M.L., N.L.S., A.L.P. and A.M.R. planned, performed and analysed the cellular immunologic and virologic assays. R.T.C., L.E.G., M.S.S., G.L., D.N.F. and D.C.M. planned, performed and analysed the humoral immunologic assays. P.A., M.J.H., M.G.P. and J.G. prepared and quality controlled the vaccine vectors. A.C. and K.G.M. led the animal work. D.H.B. designed and led the study. All coauthors discussed the data and contributed to writing the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to D.H.B. ([dbarouch@bidmc.harvard.edu](mailto:dbarouch@bidmc.harvard.edu)).



## LETTERS

# Live-animal tracking of individual haematopoietic stem/progenitor cells in their niche

Cristina Lo Celso<sup>1,4</sup>, Heather E. Fleming<sup>1,4</sup>, Juwelle W. Wu<sup>3,4</sup>, Cher X. Zhao<sup>1,4</sup>, Sam Miake-Lye<sup>1</sup>, Joji Fujisaki<sup>3,4</sup>, Daniel Côté<sup>3†</sup>, David W. Rowe<sup>5</sup>, Charles P. Lin<sup>3,4</sup> & David T. Scadden<sup>1,2,4,6</sup>

Stem cells reside in a specialized, regulatory environment termed the niche that dictates how they generate, maintain and repair tissues<sup>1,2</sup>. We have previously documented that transplanted haematopoietic stem and progenitor cell populations localize to subdomains of bone-marrow microvessels where the chemokine CXCL12 is particularly abundant<sup>3</sup>. Using a combination of high-resolution confocal microscopy and two-photon video imaging of individual haematopoietic cells in the calvarium bone marrow of living mice over time, we examine the relationship of haematopoietic stem and progenitor cells to blood vessels, osteoblasts and endosteal surface as they home and engraft in irradiated and *c-Kit*-receptor-deficient recipient mice. Osteoblasts were enmeshed in microvessels and relative positioning of stem/progenitor cells within this complex tissue was nonrandom and dynamic. Both cell autonomous and non-autonomous factors influenced primitive cell localization. Different haematopoietic cell subsets localized to distinct locations according to the stage of differentiation. When physiological challenges drove either engraftment or expansion, bone-marrow stem/progenitor cells assumed positions in close proximity to bone and osteoblasts. Our analysis permits observing in real time, at a single cell level, processes that previously have been studied only by their long-term outcome at the organismal level.

Mammalian stem cell niches have largely been defined based on localization of stem cells, with only limited definition of the heterogeneous cells within the niche that regulate stem cell function. None of these niches has been examined in a physiological context *in vivo*. The haematopoietic stem cell (HSC) niche has been studied by alteration of bone components that induce a change in stem cell function<sup>4–7</sup> or by immunohistochemistry<sup>8–10</sup>. The former does not precisely define localization and the latter does not define function. Several studies have documented the regulatory role of osteoblasts on HSC fate<sup>4,7,8,11–13</sup>; however, other studies have shown that most immunophenotypically defined HSCs can be observed adjacent to the vasculature in bone marrow<sup>9</sup>.

Bone-marrow vasculature includes arteries that penetrate compact bone at regular intervals, arborizing into capillaries that converge into a central sinus<sup>14</sup>. CD31 immunostaining of long bones enabled visualization of this vasculature network adjacent to the endosteal surface, particularly in trabecular regions where HSCs are known to localize (Fig. 1a and data not shown)<sup>7,10</sup>.

To gain a detailed, dynamic view of haematopoietic stem and progenitor cell (HSPC) localization we used intravital microscopy, scanning 4 mm × 6 mm of mouse calvarium that included the central sinus and the surrounding bone-marrow cavities<sup>3</sup> (Fig. 1b). HSC

frequencies in this region are comparable to long bones by immunophenotype and repopulating ability (Supplementary Fig. 2c, d). Because the thickness of long bones precludes analysis by intravital microscopy, we cannot confirm that our results extend to them, particularly the poorly trabeculated diaphyseal regions.

Combined confocal and two-photon microscopy permitted visualization to a depth of ~150 µm or 40–60% of the bone marrow cavity in over 75% of our measurements. Simultaneous detection of bone, osteoblasts, vasculature and labelled HSPCs was achieved by second harmonic generation (SHG) microscopy (see Methods), osteoblast-restricted collagen 1α promoter (Col2.3–GFP) reporter mice<sup>15</sup>, non-targeted quantum dots (Qdot 655 or 800) injected immediately before imaging, and flow cytometrically sorted cells stained with the lipophilic cyanine dyes Vybrant DiD (1,1'-dioctadecyl-3,3,3'-tetramethylindodicarbocyanine perchlorate) or DiI (1,1'-dilinoleyl-3,3,3',3'-tetramethylindodicarbocyanine perchlorate) (Invitrogen)<sup>3</sup>, respectively (Fig. 1d–h, Supplementary Fig. 1 and Supplementary Movie 1). Multiple combinations of cell surface markers were used to yield highly enriched long-term reconstituting HSC populations (LT-HSCs)<sup>16,17</sup>. LKS (that is, lineage<sup>low</sup> (Lin<sup>low</sup>), *c-Kit* (Kit)<sup>+</sup> and *Sca1*<sup>+</sup>) CD34<sup>+</sup>Flk2<sup>+</sup>, LKS CD150<sup>+</sup>CD48<sup>+</sup> and LKS CD48<sup>+</sup>Flk2<sup>+</sup> cells were isolated, and represent partially overlapping populations (Supplementary Fig. 7). Given that all flow cytometric methods of cell isolation can only enrich for HSCs, we will refer to all the populations used as haematopoietic stem/progenitor cells (HSPCs) to account for any potential progenitor cell contamination. The isolated cells were not adversely affected by DiD staining based on competitive and serial reconstitution assays of irradiated hosts (Supplementary Fig. 2a, b). Owing to a lack of reporters reliably marking endogenous HSPCs, transplant models were used to track the interactions of HSPCs with the microanatomy of the calvarium.

Z-stacks and three-dimensional reconstructions revealed that green fluorescent protein (GFP)-positive osteoblasts were large, irregular and relatively flat (thickness ~10 µm) cells discontinuously distributed along the endosteal surface (Fig. 1d–g). Osteoblasts were adjacent or in close proximity to vasculature (>60% within 10 µm and >90% within 20 µm of vasculature) (Fig. 1e). Whereas the middle of the cavity contained vasculature that was not associated with osteoblasts, the endosteal region demonstrated close co-localization of osteoblasts and vessels, suggesting that osteoblast-associated HSPCs would probably also be susceptible to paracrine influence exerted by vascular/perivascular cells.

The efficiency of three-dimensional *in vivo* imaging of HSPCs was compared with immunohistochemistry by injecting 375,000 to 6 × 10<sup>6</sup> Lin<sup>low</sup> DiD-labelled cells and recording the number of cells

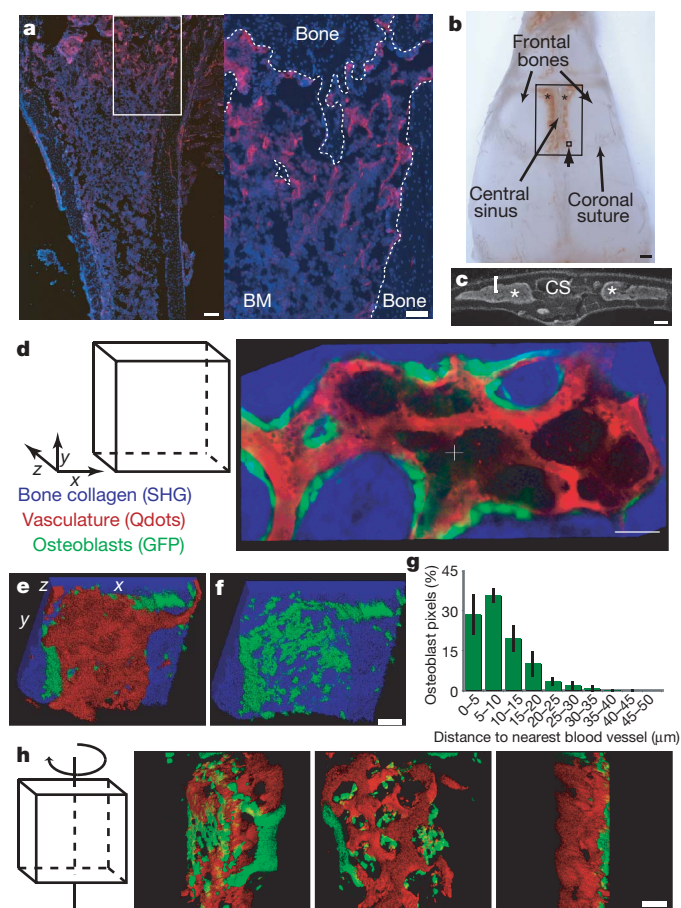
<sup>1</sup>Center for Regenerative Medicine and <sup>2</sup>Cancer Center, <sup>3</sup>Advanced Microscopy Program, Center for Systems Biology and Wellman Center for Photomedicine, Massachusetts General Hospital, 185 Cambridge Street, Boston, Massachusetts 02114, USA. <sup>4</sup>Harvard Stem Cell Institute, 42 Church Street, Cambridge, Massachusetts 02138, USA. <sup>5</sup>University of Connecticut Health Center, 663 Farmington Avenue, Farmington, Connecticut 06030, USA. <sup>6</sup>Department of Stem Cell and Regenerative Biology, Harvard University, 42 Church Street, Cambridge, Massachusetts 02138, USA. †Present address: Centre de Recherche Université Laval Robert-Giffard, Département de Physique, Université Laval, Québec, Québec G1J 2G3, Canada.

observed per *in vivo* imaged field or per 20  $\mu\text{m}$  fixed, decalcified frozen section. *In vivo* imaging was consistently more sensitive (Supplementary Fig. 3b). To define the boundaries of our *in vivo* detection system, we performed dilution experiments and noted a linear correlation between the number of DiD<sup>+</sup> imaged cells and the number of LKS CD34<sup>+</sup>Flk2<sup>+</sup> cells injected (Supplementary Fig. 3c). The numbers of cells detected when injecting  $5\text{--}15 \times 10^3$  LT-HSC-enriched populations were small (3–15 cells per recipient) but consistent, in keeping with the small fraction of the recipients' total bone marrow represented by calvarium<sup>18</sup> and the inefficiencies of HSPC homing<sup>19</sup>.

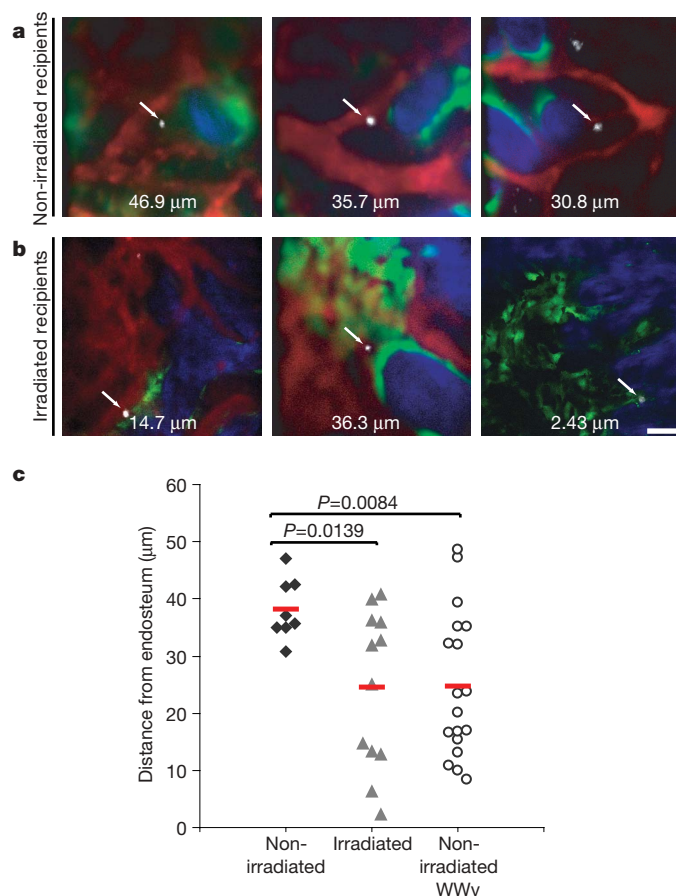
HSPCs were detected in the bone marrow cavity as early as 20 min after injection and were static throughout each 1.5- to 3-h-long imaging session. In agreement with a previous study<sup>20</sup>, cells were

dispersed across the imaged region (Supplementary Fig. 4). To assess precise anatomical relationships in the niche, we acquired Z-stack images containing the best focal plane for the DiD signal and the focal plane of the closest endosteal surface (Supplementary Fig. 5). Only HSPCs that were within 60  $\mu\text{m}$  of an endosteal surface were further analysed because beyond that distance we could not distinguish between cells located in the central cavity and those adjacent to deeper endosteum beyond our detection range (see Supplementary Movie 1). This approach excluded only 13% of all imaged cells, mostly detected at depths where SHG, GFP and Qdot signals were too faint to be collected reliably.

We adapted the established lodgement assay<sup>10,21</sup> to observe HSPC localization within the bone marrow of non-irradiated recipients (Fig. 2a). All cells were detected in close proximity to vasculature (0–16  $\mu\text{m}$ ) and at variable distances from the endosteum, but never closer than 30  $\mu\text{m}$  (Fig. 2c). Two weeks after injection, we found single, bright, DiD-positive cells located between 15–40  $\mu\text{m}$  from the endosteum (Supplementary Fig. 6a). As expected, no contribution of injected HSPCs to the host peripheral blood or bone marrow was



**Figure 1 | The calvarium endosteal niche is perivascular.** **a**, Anti-CD31 immunofluorescence (red) reveals a complex vascular network in the trabecular region of the tibia and along diaphysis endosteal surface (blue: 4,6-diamidino-2-phenylindole (DAPI) nuclear counter-stain). The right panel shows the boxed area at higher magnification and the dashed line highlights the endosteal surface. Scale bars: left, 200  $\mu\text{m}$ ; right, 100  $\mu\text{m}$ . BM, bone marrow. **b**, Intravital microscopy was used to scan the region of mouse calvarium containing bone marrow (asterisks) within the frontal bones. The large box shows the entire scanned area; the small box (arrowhead) shows one approximate observed field (330  $\mu\text{m}^2$ ). Scale bar: 1 mm. **c**, DAPI-stained calvarium coronal section indicating cavity morphology (asterisks mark the bone marrow) and depth of scanning (bracket). **d**, xy collage image of bone (blue), osteoblasts (green) and vasculature (red) acquired simultaneously with two-photon microscopy. (See also Supplementary Movie 1.) **e**, **f**, Three-dimensional reconstructions of Z-stack with all three components (**e**) or bone and osteoblasts only (**f**). **g**, Diagram representing measured distance between all osteoblast pixels and the closest vasculature ( $n = 4$  cavities; error bars indicate s.d.). **h**, Rotating three-dimensional model of osteoblasts and vasculature. Consistent results were seen with ten mice. Scale bars for **d**, **f**, **h** indicate 50  $\mu\text{m}$ .



**Figure 2 | Engrafting HSPCs reach the endosteum.** **a**, **b**, LKS CD48<sup>+</sup>Flk2<sup>+</sup> cells injected into non-irradiated (**a**,  $n = 3$  mice) or irradiated recipients (**b**,  $n = 3$  mice) were imaged within 5 h of transplantation. Two-photon microscopy was used to detect collagen bone SHG (blue) and confocal microscopy was used to detect DiD signal (white), Qdot vascular dye (red) and GFP-positive osteoblasts (green) in all panels apart from bottom right, where all signals were acquired with two-photon microscopy. Arrows point to single homed HSPCs and the numbers are the xy distance measured from each cell to the closest endosteum (edge of blue signal). Scale bar: 50  $\mu\text{m}$ . **c**, The shortest three-dimensional HSPC-endosteal surface distance was plotted for each cell imaged in non-irradiated, irradiated and WWv recipients. The average distance of HSPCs from the endosteum in irradiated and WWv recipients ( $n = 3$  mice) was significantly less than in non-irradiated recipients. Red lines represent the mean of all measurements for each set of experiments.



observed for up to 4 months after transplant (data not shown; see also ref. 22). HSPCs injected into non-irradiated hosts localized farther than 30  $\mu\text{m}$  from bone and did not engraft. The lack of engraftment of transplanted HSPCs in physiological niches imposes an inherent limitation on current studies aimed to characterize functional HSCs in their physiological niches *in vivo*.

To visualize HSPCs that could engraft, lethally irradiated wild-type and Col2.3-GFP recipients received HSPC transplants and were imaged 20 min to 5 h after injection. Irradiation compromised the vasculature: the quantum dot fluorescent signal no longer clearly demarcated bone-marrow vessels but instead spilled over the entire cavity (Fig. 2b, red), revealing progressive damage (Fig. 3a, b). Notably, circulating cells followed narrow and well defined trajectories within the quantum dot signal (data not shown), indicating that, after irradiation, bone-marrow vasculature becomes permeable to small particles such as quantum dots but not blood cells.

HSPCs injected into irradiated recipients also localized at variable distances from the endosteum, but under these conditions 47% of the cells were within 15  $\mu\text{m}$  of the endosteal surface (Fig. 2c). We monitored the mice for up to 6 months and observed at least 25% peripheral blood multilineage engraftment in all recipients, thus confirming the functional capacity of the injected cells.

WWv mice carry two mutations in the stem-cell-factor receptor *c-Kit*, resulting in an impaired ability of endogenous HSCs to engage their niche<sup>23</sup>. As expected, HSPCs injected in non-irradiated WWv mice engraft and overtake endogenous HSC production of all peripheral blood lineages<sup>23</sup> (Supplementary Fig. 6d). Whereas WWv mice show extramedullary haematopoiesis, secondary bone-marrow transplants confirmed that the injected wild-type HSPCs reside in functional, supportive bone-marrow niches (Supplementary Fig. 6e). Even in the absence of radiation-damaged vasculature, wild-type

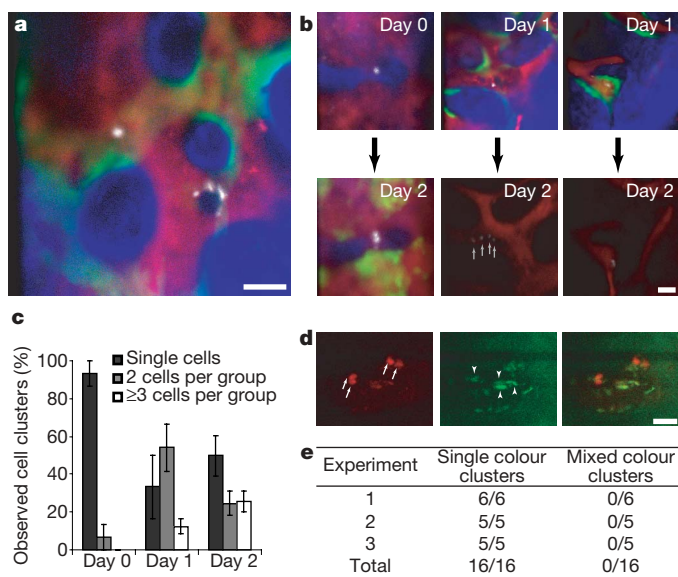
HSPCs lodged closer to bone surfaces in WWv animals than in wild-type non-irradiated mice (Fig. 2c). Therefore, multiple settings in which donor cells engraft the bone marrow are associated with HSPCs in the bone marrow in close proximity to endosteum.

Engrafting HSCs proliferate within the first 24 h after transplantation<sup>24</sup> and their resulting progeny pass through the multipotent and early progenitor states within 4 days<sup>25</sup>, but great heterogeneity in the engraftment features of stem cell populations has been previously reported<sup>26</sup>. Using our system, the *xyz* position of individual cells and their immediate progeny can be tracked *in vivo* over time (Fig. 3a, b), revealing the progressive appearance of cell clusters (Fig. 3c) of decreased dye intensity, consistent with the partitioning of DiD label on cell division (Fig. 3b). To test whether cell clusters arose from an accumulation of migrating cells, we mixed LT-HSC-enriched populations stained with DiD or a second lipophilic dye, DiI, and imaged recipient mice 2 and 3 days after injection. Only a single dye was observed in cells within each cluster (Fig. 3d, e), consistent with the proliferation of single cells associated with early engraftment. Moreover, we observed 5-bromodeoxyuridine (BrdU) uptake by the injected cells using both immunofluorescence and fluorescence-activated cell sorting (FACS) analysis (Supplementary Fig. 9). Our analysis revealed heterogeneity of cell division patterns at the single cell level *in vivo*.

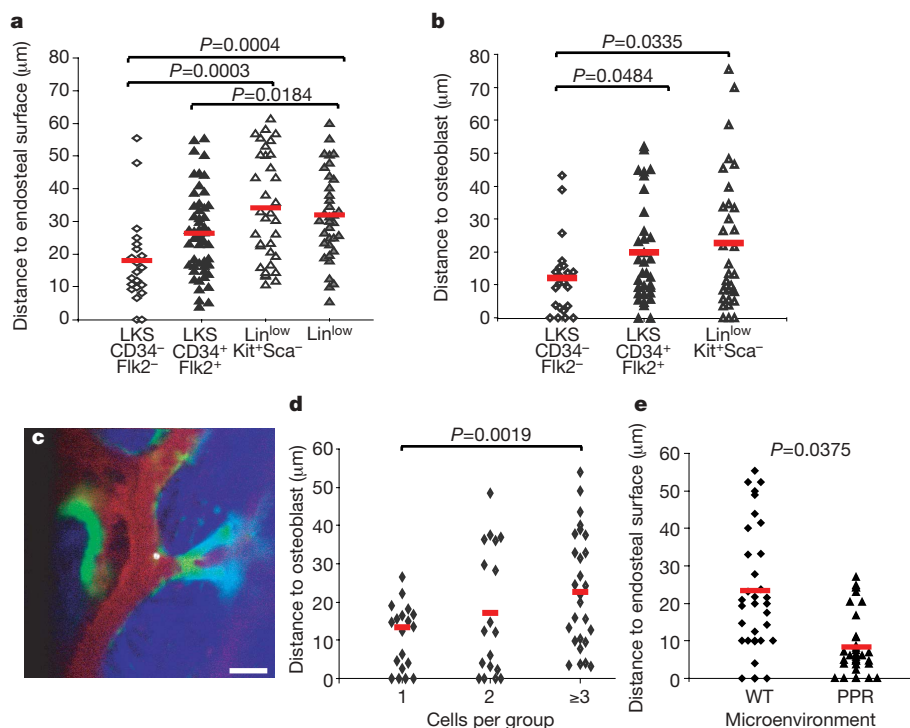
We sought to determine whether positioning within the bone-marrow microenvironment was influenced by intrinsic features of the haematopoietic cells such as differentiation state. We imaged LT-HSC-enriched (LKS CD34<sup>+</sup>Flk2<sup>+</sup>), multipotent-progenitor-enriched (MPP; LKS CD34<sup>+</sup>Flk2<sup>+</sup>)<sup>27,28</sup> and committed-progenitor-enriched (Lin<sup>low</sup>Kit<sup>+</sup>Sca<sup>+</sup>) populations and the heterogeneous Lin<sup>low</sup> cells within hours after transplantation (Fig. 4a, b). LT-HSC-enriched cells localized closest to endosteum and osteoblasts, with more mature subsets residing progressively farther away. To determine whether cell position correlated with cell division, we compared the position of cells that had proliferated creating clusters of  $\geq 3$  cells with those that did not divide. The quiescent cells were significantly closer to osteoblasts (Fig. 4d).

To determine whether niche variables could influence HSPC localization, we injected cells in wild-type or PPR mice (that is, mice transgenic for a constitutively active parathyroid hormone/parathyroid hormone related peptide receptor driven by the osteoblast specific promoter, *Col12.3kb*)<sup>29</sup>. PPR mice have increased osteoblast number, trabecular bone volume and HSCs<sup>4</sup>, and have been shown to drive expansion of injected HSCs when recipients of bone-marrow transplantation<sup>30</sup>. Two days after injection, DiD<sup>+</sup> cells in PPR mice were markedly closer to the endosteum even when correcting for the reduced bone-marrow cavity size (Fig. 4e and data not shown). Therefore, stem cell non-autonomous features of the niche contribute to the regulation of HSPC localization.

The data presented here demonstrate that the previously proposed dichotomy between distinct osteoblast and perivascular niches is not anatomically feasible in the calvarium: osteoblasts are perivascular. We cannot exclude perivascular-only niches within the diaphysis of long bones or in non-bony tissues such as spleen; however, the microarchitecture of trabeculae indicates a joint periendosteal-perivascular niche. The relationship of transplanted primitive haematopoietic cells to anatomical components of the trabecular niche is influenced by cell-intrinsic and niche-intrinsic variables. Cells of differing immunophenotype localized differently, with closer proximity to bone corresponding to a greater enrichment for stem cell function. In addition, conditions in the microenvironment that enable stem cell engraftment or in which osteoblasts drive HSC expansion yield closer association of HSPCs with bone. There is anatomical dynamism within the bone marrow, with cell positioning reflecting differing physiological demands on stem cells. However, we cannot conclude that HSPC engraftment or expansion require direct osteoblast contact, as this was not uniformly observed. Rather, approximate vicinity may be sufficient for osteoblast alteration of



**Figure 3 | Engraftment is initiated by asynchronous HSPC cell divisions.** **a**, HSPC progeny were imaged 1 day after injection in irradiated recipients ( $n = 4$  mice), revealing heterogeneity in cell clustering. Blue, bone; red, vasculature; green, osteoblasts; white, HSPC progeny. **b**, Cells were tracked from day 0 to day 2 ( $n = 2$  recipient mice) or from day 1 to day 2 ( $n = 3$  recipient mice) and diverse kinetics of cell division were observed. **c**, Increasing numbers of clusters containing 2 or  $\geq 3$  cells were observed in the days after injection ( $n = 4$ ; error bars indicate s.e.m.). **d**, When 50% of cells were stained with DiD and 50% with DiI before injection, only single-colour clusters were observed. Red, DiD; green, autofluorescence. Arrows point at each DiD-positive cell within two clusters; arrowheads point at autofluorescent cells. Cells accepted for assessment had a dye/autofluorescence signal ratio  $> 2$  (in this example 2.82, 3.71, 4.12, 8.22). The same analysis was used to validate DiI signals. Scale bars in **a**, **b**, **d** are 50  $\mu\text{m}$ . **e**, Summary of observed cell clusters in three independent experiments.



**Figure 4 | Cell-dependent and niche-dependent HSPC localization.** **a, b,** LKS CD34<sup>-</sup> Flk2<sup>-</sup> ( $n = 6$  recipients) LT-HSC-enriched, LKS CD34<sup>+</sup> Flk2<sup>+</sup> MPP ( $n = 3$  recipients), Lin<sup>low</sup> Kit<sup>+</sup> Sca<sup>-</sup> progenitors ( $n = 3$  recipients) and Lin<sup>low</sup> ( $n = 4$  recipients) populations home to distinct locations, closer to or further from endosteum (**a**) and osteoblasts (**b**). All imaging was performed within 5 h of transplantation. **c,** Representative image of a LKS CD34<sup>-</sup> Flk2<sup>-</sup> cell residing adjacent to an osteoblast and close to endosteum 4 h after injection. White, DiD-labelled cell; green, GFP<sup>+</sup> osteoblasts; red, Qdot vascular dye;

blue, bone collagen. Scale bar: 50 μm. **d,** Independently of the LT-HSC-enriched starting population, the distance between DiD-labelled cells and osteoblasts measured 2 days after injection increased with the number of cells found in each observed cluster ( $n = 5$  imaged mice). **e,** Two days after injecting the same number of LKS cells in irradiated wild-type or PPR mice ( $n = 2 + 2$ ) similar numbers of cells were observed, but LKS progeny in PPR mice were preferentially located adjacent to the endosteal surface. Red lines indicate the mean of all measurements in each set of experiments.

stem cell function. Gradients of secreted factors and extracellular matrix, or events induced in other surrounding cell types, may contribute to the regulatory function of osteoblasts in the niche.

We have visualized single transplanted HSPCs in their niche within a living mammal and demonstrated that the state of both haematopoietic cell and microenvironment affects the physical association between components of haematopoietic tissue. Continued technical developments and the generation of more sophisticated molecularly modified animal models will permit further analysis of physiologically relevant settings—other than transplantation or artificially permissive environments—and of other stem cells of normal and malignant types as they engage and perhaps compete for niche elements. The combination of molecular, cellular and organismal biology may thereby be simultaneously examined to understand the regulatory networks affecting stem cell control.

## METHODS SUMMARY

All mice were housed according to IACUC guidelines and used for experiment when 8–14-weeks old. Wild-type C57BL/6.SJL mice were HSPC donors when recipients were wild type, Col2.3–GFP or WWv double mutant (backcrossed to C57BL/6 background). FVB mice were donors for wild-type or PPR littermate mice<sup>4,29</sup> (gift from E. Schipani).

Mice were anaesthetized and prepared for *in vivo* imaging as described<sup>3</sup>. Immediately before imaging 20 μl of non-targeted Qdot 800 or 655 (Invitrogen) diluted in 130 μl sterile PBS was injected retro-orbitally to allow vasculature visualization. The mouse was held in a heated tube mounted on a precision 3 axis motorized stage (Suter MP385). All mice were imaged with a custom-built confocal two-photon hybrid microscope specifically designed for live animal imaging (see Methods). At the start of each imaging session, we surveyed large areas of the skull bone surface using video rate second harmonic microscopy (see Methods) to identify the major anatomical landmarks such as sagittal and coronal sutures. We identified the locations of HSPCs within bone-marrow cavities and recorded their coordinates relative to the intersection of the

sagittal and coronal sutures. SHG and GFP signals above each identified HSPC were acquired every 5 to 20 μm until the above endosteal surface was reached. After *in vivo* imaging, the scalp was re-closed using 3 M Vetbond veterinary glue and post-operative care was provided as described<sup>3</sup>.

Images were coloured and merged using Adobe Photoshop and HSPC-microenvironment distance measures were obtained using Adobe Illustrator and Microsoft Excel. A two-tailed type 2 *t*-test was applied to all data. *P* values ≤0.05 were considered statistically significant.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 23 April; accepted 17 September 2008.

Published online 3 December 2008.

- Schofield, R. The relationship between the spleen colony-forming cell and the haematopoietic stem cell. *Blood Cells* **4**, 7–25 (1978).
- Xie, T. & Spradling, A. C. A niche maintaining germ line stem cells in the *Drosophila* ovary. *Science* **290**, 328–330 (2000).
- Sipkins, D. A. *et al.* *In vivo* imaging of specialized bone marrow endothelial microdomains for tumour engraftment. *Nature* **435**, 969–973 (2005).
- Calvi, L. M. *et al.* Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature* **425**, 841–846 (2003).
- Nilsson, S. K. *et al.* Osteopontin, a key component of the hematopoietic stem cell niche and regulator of primitive hematopoietic progenitor cells. *Blood* **106**, 1232–1239 (2005).
- Stier, S. *et al.* Osteopontin is a hematopoietic stem cell niche component that negatively regulates stem cell pool size. *J. Exp. Med.* **201**, 1781–1791 (2005).
- Zhang, J. *et al.* Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* **425**, 836–841 (2003).
- Arai, F. *et al.* Tie2/angiopoietin-1 signaling regulates hematopoietic stem cell quiescence in the bone marrow niche. *Cell* **118**, 149–161 (2004).
- Kiel, M. J. *et al.* SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121**, 1109–1121 (2005).
- Nilsson, S. K., Johnston, H. M. & Coverdale, J. A. Spatial localization of transplanted hemopoietic stem cells: inferences for the localization of stem cell niches. *Blood* **97**, 2293–2299 (2001).



11. Lord, B. I., Testa, N. G. & Hendry, J. H. The relative spatial distributions of CFUs and CFUc in the normal mouse femur. *Blood* **46**, 65–72 (1975).
12. Taichman, R. S., Reilly, M. J. & Emerson, S. G. The Hematopoietic microenvironment: Osteoblasts and the hematopoietic microenvironment. *Hematology* **4**, 421–426 (2000).
13. Mayack, S. R. & Wagers, A. J. Osteolineage niche cells initiate hematopoietic stem cell mobilization. *Blood* **112**, 519–531 (2008).
14. Crock, H. V. *The Blood Supply of the Lower Limb Bones in Man* (E&S Livingstone LTD, 1967).
15. Kalajzic, Z. *et al.* Directing the expression of a green fluorescent protein transgene in differentiated osteoblasts: comparison between rat type I collagen and rat osteocalcin promoters. *Bone* **31**, 654–660 (2002).
16. Bryder, D., Rossi, D. J. & Weissman, I. L. Hematopoietic stem cells: the paradigmatic tissue-specific stem cell. *Am. J. Pathol.* **169**, 338–346 (2006).
17. Kiel, M. J. *et al.* Haematopoietic stem cells do not asymmetrically segregate chromosomes or retain BrdU. *Nature* **449**, 238–242 (2007).
18. Abkowitz, J. L., Catlin, S. N., McCallie, M. T. & Guttorm, P. Evidence that the number of hematopoietic stem cells per animal is conserved in mammals. *Blood* **100**, 2665–2667 (2002).
19. Lewin, M. *et al.* Tat peptide-derivatized magnetic nanoparticles allow *in vivo* tracking and recovery of progenitor cells. *Nature Biotechnol.* **18**, 410–414 (2000).
20. Suzuki, N. *et al.* Combinatorial Gata2 and Sca1 expression defines hematopoietic stem cells in the bone marrow niche. *Proc. Natl Acad. Sci. USA* **103**, 2202–2207 (2006).
21. Adams, G. B. *et al.* Stem cell engraftment at the endosteal niche is specified by the calcium-sensing receptor. *Nature* **439**, 599–603 (2006).
22. Zhong, J. F., Zhan, Y., Anderson, W. F. & Zhao, Y. Murine hematopoietic stem cell distribution and proliferation in ablated and nonablated bone marrow transplantation. *Blood* **100**, 3521–3526 (2002).
23. Migliaccio, A. R., Carta, C. & Migliaccio, G. *In vivo* expansion of purified hematopoietic stem cells transplanted in nonablated W/W<sup>v</sup> mice. *Exp. Hematol.* **27**, 1655–1666 (1999).
24. Nilsson, S. K., Dooner, M. S. & Quesenberry, P. J. Synchronized cell-cycle induction of engrafting long-term repopulating stem cells. *Blood* **90**, 4646–4650 (1997).
25. Forsberg, E. C., Serwold, T., Kogan, S., Weissman, I. L. & Passegue, E. New evidence supporting megakaryocyte-erythrocyte potential of flk2/flt3<sup>+</sup> multipotent hematopoietic progenitors. *Cell* **126**, 415–426 (2006).
26. Dykstra, B. *et al.* Long-term propagation of distinct hematopoietic differentiation programs *in vivo*. *Cell Stem Cell* **1**, 218–229 (2007).
27. Adolfsson, J. *et al.* Identification of Flt3<sup>+</sup> lympho-myeloid stem cells lacking erythro-megakaryocytic potential: a revised road map for adult blood lineage commitment. *Cell* **121**, 295–306 (2005).
28. Osawa, M., Hanada, K., Hamada, H. & Nakauchi, H. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science* **273**, 242–245 (1996).
29. Calvi, L. M. *et al.* Activated parathyroid hormone/parathyroid hormone-related protein receptor in osteoblastic cells differentially affects cortical and trabecular bone. *J. Clin. Invest.* **107**, 277–286 (2001).
30. Adams, G. B. *et al.* Therapeutic targeting of a stem cell niche. *Nature Biotechnol.* **25**, 238–243 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank E. Schipani for providing the PPR mice. We are grateful for help and advice from A. Catic, L. Purton, V. Janzen, G. Adams, J. Spencer, J. Runnels and P. O'Donovan. We thank Y. Tang for the mice husbandry care; D. Dombkowski, L. Prickett and K. Folz-Donahue for cell sorting expertise; R. Klein and K. Chomsky-Higgins for technical assistance; and C. Pasker, V. Shannon, M. Indico Miklosik and D. Machon for administrative assistance. C.L.C. was funded by EMBO and HFSP. The project was funded by the National Institutes of Health (to D.T.S. and C.P.L.), the Harvard Stem Cell Institute (to C.P.L.) and philanthropic sources (to D.T.S. and C.L.C.).

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details accompany the full-text HTML version of the paper at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to D.T.S. ([dscadden@mgh.harvard.edu](mailto:dscadden@mgh.harvard.edu)) or C.P.L. ([lin@helix.mgh.harvard.edu](mailto:lin@helix.mgh.harvard.edu)).

## METHODS

**Microscopy.** Second harmonic signal is generated by collagen in the bone when it is illuminated by femtosecond titanium:sapphire laser pulses<sup>31</sup>. The exact wavelength for SHG was not critical as it is a scattering rather than an absorption process. GFP-expressing osteoblasts were imaged by either two-photon (920 nm excitation, 500–590 nm detection) or confocal microscopy (491 nm excitation, 505–590 nm detection). Circulating Qdot 655 was imaged with two-photon microscopy (920 nm excitation, 625–675 nm detection). Qdot 800 was imaged with confocal microscopy (532 nm excitation, >795 nm detection). DiD-labelled cells were identified and distinguished from autofluorescent cells by taking two confocal images at 633 nm (650–760 nm detection) and at 532 nm (560–640 nm detection), respectively. The opposite settings were used to identify DiI-labelled cells.

A  $\times 30$  0.9NA water immersion objective (Lomo) was used for all imaging, so that each field observed always corresponded to  $330 \times 330 \mu\text{m}$ . To minimize chromatic aberration owing to the wide range of excitation wavelengths used, the collimation of each excitation laser beam was adjusted with a telescope to achieve parfocality. For three-dimensional analysis of bone-marrow cavity, Z-stacks were acquired at 1–3  $\mu\text{m}$  steps. A PCI-based image capture board (Snapper24, Active Silicon) was used to acquire up to three channels simultaneously using iPhon<sup>32</sup> software running under Mac OS X that was developed in-house.

**Image processing.** Each channel was acquired in 8-bit greyscale and merged to an RGB image using iPhon or Photoshop (Qdot, red; GFP, green; collagen, blue). The fourth channel (DiD) was given a white colour code and merged with the RGB image. The intensity for each channel was scaled so that 0–255 in the 8-bit image corresponded to  $7 - I_{\text{max}}$  in the original image, where 7 was the noise floor and  $I_{\text{max}}$  was the maximum pixel value.

Two-dimensional measurements were performed by drawing and measuring lines using Adobe Illustrator and importing images at  $500 \times 500$  point size. Points-to-micrometres conversion and all data analysis was performed using Microsoft Excel software.

Three-dimensional rendering was obtained using ImageJ 1.38 (open source NIH software, <http://rsb.info.nih.gov/ij/index.html>). 8-bit stacks from each channel were given a threshold to eliminate hue differences within structures and assigned an arbitrary intensity (highest to bone, lowest to vasculature and intermediates to HSCs and osteoblasts) and combined in opaque models. Bright and dark outlier pixels were eliminated to reduce noise and each structure was assigned blue/red/green/white colour from the LUT table. The VolumeJ plugin was used for 'cine-mode' rendering with the following settings: classifier ramp + index, scaling  $\geq 1$ , aspect ratio 1:1:1 or 1:1:3 depending on the Z-stack step, light 0, 0, –1,000.

To measure distances between osteoblasts and vasculature, multiple image stacks from the same bone cavity were concatenated based on coordinate information. The resultant stack was imported into ImageJ, re-sliced into the vertical planes (Y–Z, X–Z) and digitally rotated to remove skull tilt relative to imaging plane. The stack channels were separated, and the greyscale intensity profile along the Z direction of the bone was measured and fit into an exponential decay curve as a function of depth, the inverse of which was then multiplied to the bone (blue) channel. A similar function was applied to the osteoblast (green) and the vascular (red) channels using an empirically derived exponent. A multiplicative constant was incorporated to equalize channel brightness. Euclidean distance maps (EDM), codifying the minimal pixel distance between each non-vascularized point to the vessel walls as a greyscale value, were generated in ImageJ based on the binarized vascular signal for the three Cartesian planes (X–Y, Y–Z and X–Z); the final three-dimensional-EDM stack was created after re-slicing the two vertical (YZ, XZ) EDMs in the XY orientation and assigning to each pixel the smallest greyscale value out of the three planar EDMs. The binary signal of the cells was inverted and the larger of the greyscale values between this stack and the vascular EDM was elected for each pixel; the resultant stack thus codified the distance from all points of the cells to the vessel wall with a consistent background value of 255. The histogram of these stacks constituted the distance distribution data.

**FACS analysis.** To analyse LT-HSC subpopulation overlap, whole bone-marrow mononuclear cells were stained with c-Kit allophycocyanine, Flk2 phycoerythrin (PE), Lineage cocktail biotin (B220, Mac1, Ter119, CD3, CD4, CD8 at 1:1:1:1:1:1) followed by streptavidin PrcP (BD Pharmingen), Sca-1 PE–Cy5.5 (Caltag), CD34 fluorescein isothiocyanate (FITC) (eBioscience), CD48 Pacific

blue and CD150 PE–Cy7 (BioLegend). At least  $2 \times 10^6$  events were acquired using a BD LSRII flow-cytometer. Compensation and data analysis were performed using Flowjo 8.5.3. BrdU was detected using the BD BrdU flow detection kit on either whole bone-marrow or lineage-depleted cells.

**HSPC isolation, staining and injection.** Whole bone-marrow cells were collected by crushing tibias, femurs, hips and spine of donor mice, stained with biotin-conjugated lineage cocktail antibodies and subjected to lineage depletion using Miltenyi magnetic beads and columns. The resulting lineage-depleted fraction was stained with the antibodies listed above and sorted using FACS DiVa or FACS ARIA (Becton Dickinson). A small fraction of the collected cells was re-run through the sorter and over 95% purity was consistently confirmed.

$1 - 5 \times 10^5$  cells per ml HSPCs were stained with 5  $\mu\text{M}$  DiD or 7.5  $\mu\text{M}$  DiI in PBS without serum for 10 min at 37 °C, washed once in PBS and immediately injected into the tail vein of recipient mice. Unless stated differently, each imaged CD45.2 mouse received 8,000 to 15,000 labelled CD45.1 HSPCs together with  $3 - 5 \times 10^5$  CD45.2 supportive whole bone-marrow mononuclear cells to ensure survival.

**Bone-marrow transplantation.** Unless otherwise stated, recipient mice were lethally irradiated with 9.5 Gy the night before transplantation. To evaluate DiD toxicity,  $10^6$  CD45.1 whole bone-marrow mononuclear cells stained or unstained with DiD were injected together with  $10^6$  CD45.2 unstained whole bone-marrow mononuclear cells into CD45.2 recipients (five mice in each group). For serial transplantation, 30,000 DiD-stained CD45.2 LKS CD48<sup>+</sup> Flk2<sup>+</sup> LT-HSCs were injected into two CD45.1 recipients. After 6 weeks, primary recipients' whole bone-marrow mononuclear cells were collected and pooled.  $2 \times 10^6$  cells were injected in each of five secondary CD45.1 recipients. Recipients' peripheral blood was assessed for CD45.1/CD45.2 chimaerism at 3, 12, 16 and 27 weeks after transplant using PE- and FITC-conjugated antibodies, respectively (BD Pharmingen). PB chimaerism was monitored also in all imaged mice.

Calvarium and femur whole bone-marrow long-term reconstitution ability was compared by competitive transplant of  $10^6$  CD45.1 calvarium or femur whole bone-marrow mononuclear cells together with  $10^6$  CD45.2 femur whole bone-marrow mononuclear cells into CD45.2 recipients. Recipients' peripheral blood chimaerism was assessed 4, 8, 12 and 16 weeks after transplantation.

The ability of HSCs to engraft fully in the bone marrow of WWv mice was checked by serial transplantation. A total of 10,000 LT-HSC-enriched cell populations were injected in non-irradiated WWv recipients and 6 months later  $2 \times 10^6$  whole bone-marrow mononuclear cells from reconstituted femurs were injected into separate CD45.2 recipients. Peripheral blood chimaerism was assessed 12 weeks later.

**BrdU administration.** Wild-type CD45.1 mice transplanted with wild-type CD45.2 LKS CD34<sup>+</sup> Flk2<sup>+</sup> or Rosa26 rtTA  $\times$  TRE H2BGFP<sup>33</sup> LKS CD150<sup>+</sup> CD48<sup>+</sup> cells received 100  $\mu\text{l}$  BrdU solution (10 mg ml<sup>−1</sup>) via intraperitoneal injection every 12 h from 12 h after transplant. Negative control mice were injected with saline. Two days after transplantation bones were harvested and either crushed for flow cytometry detection of BrdU uptake or processed for immunofluorescence.

**Immunofluorescence.** Frozen sections from paraformaldehyde fixed and decalcified tibias from wild-type mice were stained with anti-CD31 biotin-conjugated antibody (BD Pharmingen) using the Perkin Elmer Tyramide Kit following the manufacturer's instructions. Streptavidin HRP and tyramide TMR were used to visualize CD31 localization. Vectashield (VectorLabs) containing DAPI nuclear counter-stain was used to mount the sections. Images were acquired with a Nikon Eclipse 80i epifluorescence microscope equipped with a Qimaging Micropublisher digital CCD colour camera. BrdU staining was performed using the BD BrdU *in situ* detection kit according to the manufacturer's instructions and terminating with Perkin Elmer TMR tyramide amplification. Rabbit polyclonal anti-GFP antibody (Invitrogen) was incubated at the same time as the BrdU antibody and donkey anti-rabbit Alexa 647 conjugated antibody (Invitrogen) was incubated at the same time as streptavidin HRP. Images were acquired with a Bio Rad MRC600 confocal microscope.

31. Zipfel, W. R., Williams, R. M. & Webb, W. W. Nonlinear magic: multiphoton microscopy in the biosciences. *Nature Biotechnol.* 21, 1369–1377 (2003).

32. Veilleux, I., Spencer, J. A., Biss, D. P., Cote, D. & Lin, C. P. *In vivo* cell tracking with video rate multimodality laser scanning microscopy. *IEEE ISTQE* 14, 10–18 (2008).

33. Brennand, K., Huangfu, D. & Melton, D. All beta cells contribute equally to islet growth and maintenance. *PLoS Biol.* 5, e163 (2007).



# Detection of functional haematopoietic stem cell niche using real-time imaging

Yucai Xie<sup>1,2\*</sup>, Tong Yin<sup>1\*</sup>, Winfried Wiegraabe<sup>1</sup>, Xi C. He<sup>1</sup>, Diana Miller<sup>3</sup>, Danny Stark<sup>1</sup>, Katherine Perko<sup>1</sup>, Richard Alexander<sup>1</sup>, Joel Schwartz<sup>1</sup>, Justin C. Grindley<sup>1</sup>, Jungeun Park<sup>1</sup>, Jeff S. Haug<sup>1</sup>, Joshua P. Wunderlich<sup>1</sup>, Hua Li<sup>1</sup>, Simon Zhang<sup>1</sup>, Teri Johnson<sup>1</sup>, Ricardo A. Feldman<sup>3</sup> & Linheng Li<sup>1,4</sup>

Haematopoietic stem cell (HSC) niches, although proposed decades ago<sup>1</sup>, have only recently been identified as separate osteoblastic and vascular microenvironments<sup>2–6</sup>. Their interrelationships and interactions with HSCs *in vivo* remain largely unknown. Here we report the use of a newly developed *ex vivo* real-time imaging technology and immunoassaying to trace the homing of purified green-fluorescent-protein-expressing (GFP<sup>+</sup>) HSCs. We found that transplanted HSCs tended to home to the endosteum (an inner bone surface) in irradiated mice, but were randomly distributed and unstable in non-irradiated mice. Moreover, GFP<sup>+</sup> HSCs were more frequently detected in the trabecular bone area compared with compact bone area, and this was validated by live imaging bioluminescence driven by the stem-cell-leukaemia (*Scl*) promoter-enhancer<sup>7</sup>. HSCs home to bone marrow through the vascular system. We found that the endosteum is well vascularized and that vasculature is frequently localized near N-cadherin<sup>+</sup> pre-osteoblastic cells, a known niche component. By monitoring individual HSC behaviour using real-time imaging, we found that a portion of the homed HSCs underwent active division in the irradiated mice, coinciding with their expansion as measured by flow assay. Thus, in contrast to central marrow, the endosteum formed a special zone, which normally maintains HSCs but promotes their expansion in response to bone marrow damage.

To study the functional interaction between the HSC and its niche, we developed a method of *ex vivo* imaging stem cells (EVISC) (Fig. 1a and Supplementary Fig. 1). This method combines real-time imaging technology<sup>8</sup> (Supplementary Movie 1) and the ability of HSCs to home to their niche after transplantation into irradiated mice<sup>9,10</sup> (Supplementary Methods). To identify and track the transplanted HSCs, we isolated Flk2<sup>−</sup>Lin<sup>−</sup>Sca-1<sup>+</sup>c-Kit<sup>+</sup> (Flk2<sup>−</sup>LSK) cells that are enriched with purified HSCs<sup>9</sup> from actin-driven GFP-transgenic mice (Supplementary Fig. 1a, b). We compared the behaviour of these cells after transplantation into irradiated and non-irradiated mice using two-photon microscopy<sup>11,12</sup> (Fig. 1a, b). It has been documented that HSCs migrate from tail vein to bone marrow in a few hours<sup>13,14</sup>. We obtained real-time images of HSC arrival in BM around 5–8 h after transplantation. In non-irradiated mice, we detected GFP<sup>+</sup> HSCs in bone marrow (Fig. 1b and Supplementary Movie 2). In irradiated mice, however, we more frequently observed GFP<sup>+</sup> HSCs homing to the endosteal region (Fig. 1c and Supplementary Movie 3). Our real-time imaging method enabled observation of dynamic interactions between HSCs and their niches.

As the EVISC assay targeted limited areas, we obtained a more complete picture of the distribution of the transplanted GFP<sup>+</sup> HSCs

by immunostaining multiple longitudinal sections of femurs and tibias from three experiments. Positions of GFP<sup>+</sup> HSCs were measured as cell distance (where 1 cell distance (CD) = 6–8  $\mu$ m) from the endosteal surface. (Fig. 1d, e). We generated a cumulative percentage plot to display distribution patterns under irradiated versus non-irradiated conditions (Fig. 1f). In non-irradiated recipients, GFP<sup>+</sup> HSCs accumulated steadily with increasing distance from the endosteum. In contrast, in irradiated recipients, the GFP<sup>+</sup> HSCs accumulated rapidly close to the endosteum, but more gradually from 3 CD outwards. Two-sided Fisher's exact test confirmed that the cells were significantly more likely to be located within  $\leq 2$  CD under irradiated than non-irradiated conditions (58.1% versus 12.5% respectively,  $P = 0.009$ , Fig. 1g and Supplementary Table 1). Cell distribution patterns 3 CD from the endosteum and beyond were not significantly different between the two conditions ( $P = 0.9576$ ). In addition, we observed substantial reduction in bone marrow cellularity and enlarged blood vessels in irradiated bone marrow (Fig. 1d, e). Thus, after bone marrow damage, HSCs showed biased homing favouring the 'endosteal zone' ( $\leq 2$  CD) rather than central marrow ( $\geq 3$  CD).

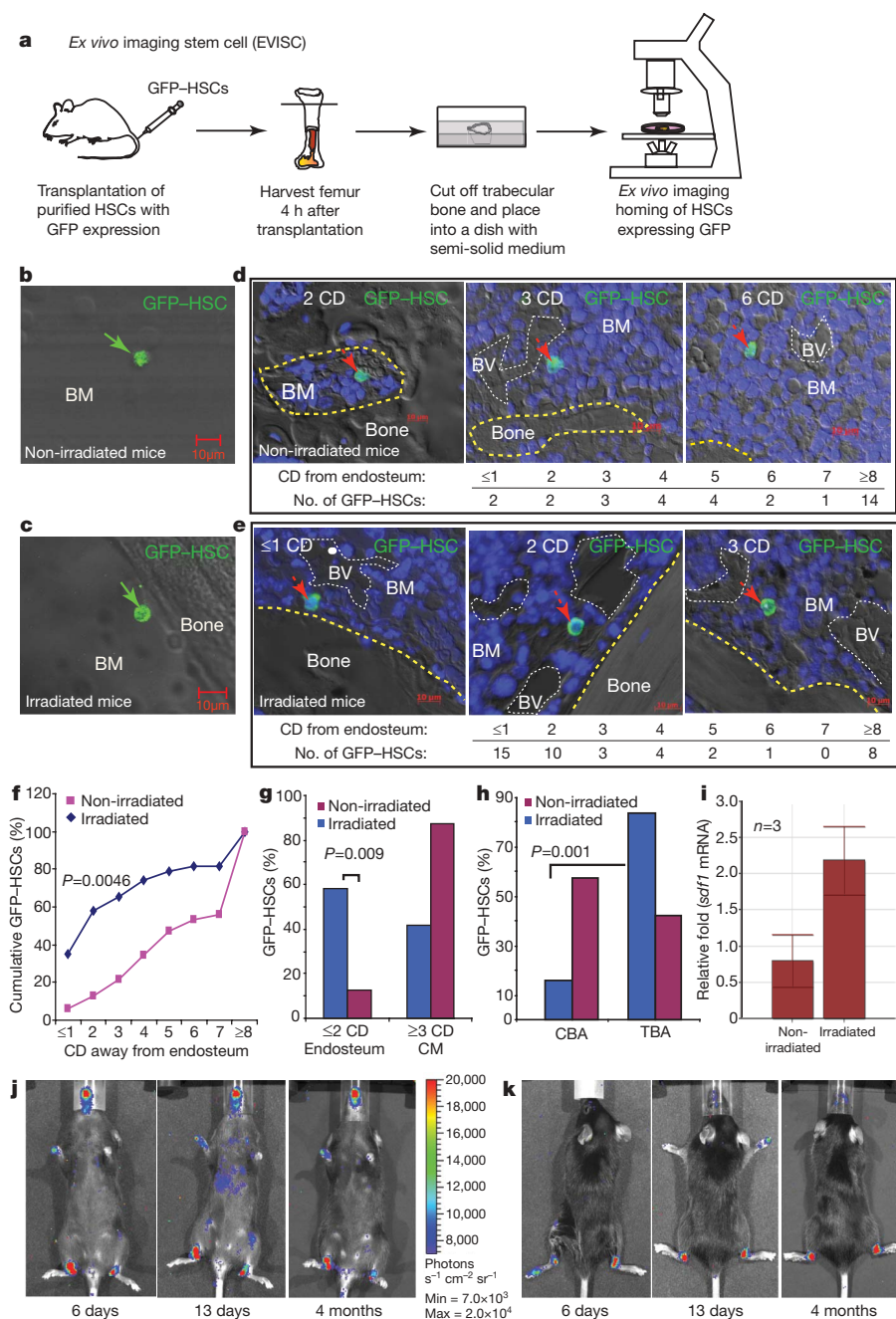
We also observed that GFP<sup>+</sup> HSCs homed predominantly to the trabecular bone area (83.8%) compared with the compact bone area (16.2%) ( $P = 0.001$ ) under irradiated conditions, but showed no preference to either trabecular or compact bone area (42.3% compared with 57.7%) under non-irradiated conditions (Fig. 1h and Supplementary Table 2). We wondered what would be the mechanism underlying the biased HSC homing, and found that expression of Sdf1 (ref. 15), a key chemotactic factor, was increased (2.7-fold) in trabecular bone area in response to irradiation (Fig. 1i).

To confirm the location of HSCs in live mice, we used *Scl*-TVA transgenic mice in which an avian retrovirus receptor is driven from *Scl*-promoter-3'enhancer regulatory elements that are active predominantly, but not exclusively, in HSCs<sup>16,17</sup>. When these mice were injected with an avian virus containing a luciferase reporter (RCAS-Luc), only the *Scl*-TVA<sup>+</sup> cells were susceptible to infection, allowing this population to be visualized by live-imaging bioluminescence, which reflected the luciferase activity<sup>7,18</sup> (Supplementary Fig. 2). Bioluminescence imaging displayed only transient signals in liver and spleen (comparing 13 days with 4 months in Fig. 1j), reflecting short-term mobilization and potential expansion of HSCs in response to bone marrow damage, but showed strong and persistent signals (>10 months) in the trabecular bone area of both legs and other regions (Fig. 1j, k and Supplementary Fig. 2c), reflecting self-renewing HSCs.

The tendency of HSCs to home to the endosteum suggested preferential homing to the osteoblastic niche, which was thought to be

<sup>1</sup>Stowers Institute for Medical Research, 1000 E. 50th Street, Kansas City, Missouri 64110, USA. <sup>2</sup>Department of Cardiology, Shanghai Rui Jin Hospital, Shanghai Jiao Tong University School of Medicine, 197, Rui Jin 2 Road, Shanghai 200025, China. <sup>3</sup>Department of Microbiology and Immunology, and Greenebaum Cancer Center, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA. <sup>4</sup>Department of Pathology and Laboratory Medicine, Kansas University Medical Center, 3901 Rainbow Boulevard, Kansas City, Kansas 66160, USA.

\*These authors contributed equally to this work.



**Figure 1 | Detection of HSCs using EVISC, immunoassaying and the *Scf*-TVA animal model.**

Differential interference contrast (DIC) images show bone structure. Diamidino-2-phenylindole (DAPI) staining is blue. BM, bone marrow; BV, blood vessel. **a**, A diagram showing the process of EVISC. **b**, **c**, Detection of a GFP<sup>+</sup> HSC (GFP-HSC) using EVISC. **d**, **e**, Representative images of GFP<sup>+</sup> HSCs (red arrow) at different positions. The average bone marrow cell diameter (6–8 μm) was used as one cell distance (CD). **f**, Cumulative percentage plot of GFP<sup>+</sup> HSC numbers at each position away from endosteum (**d**, **e**). **g**, **h**, GFP<sup>+</sup> HSC distribution in endosteal and central marrow zones (**g**) and in trabecular bone area (TBA) and compact bone area (CBA) (**h**). The Fisher's exact test was used for the statistical analysis (Supplementary Tables 1 and 2). CM, central marrow. **i**, Messenger RNA level of *Scf1* (also known as *Cxcl12*) in bone-marrow-depleted trabecular bone area ( $n=3$ , error bars represent standard deviation, s.d.). **j**, **k**, Live bioluminescence imaging of *Scf*-TVA animals injected with RCAS-Luc virus; ventral (**j**) and dorsal (**k**) views are shown.

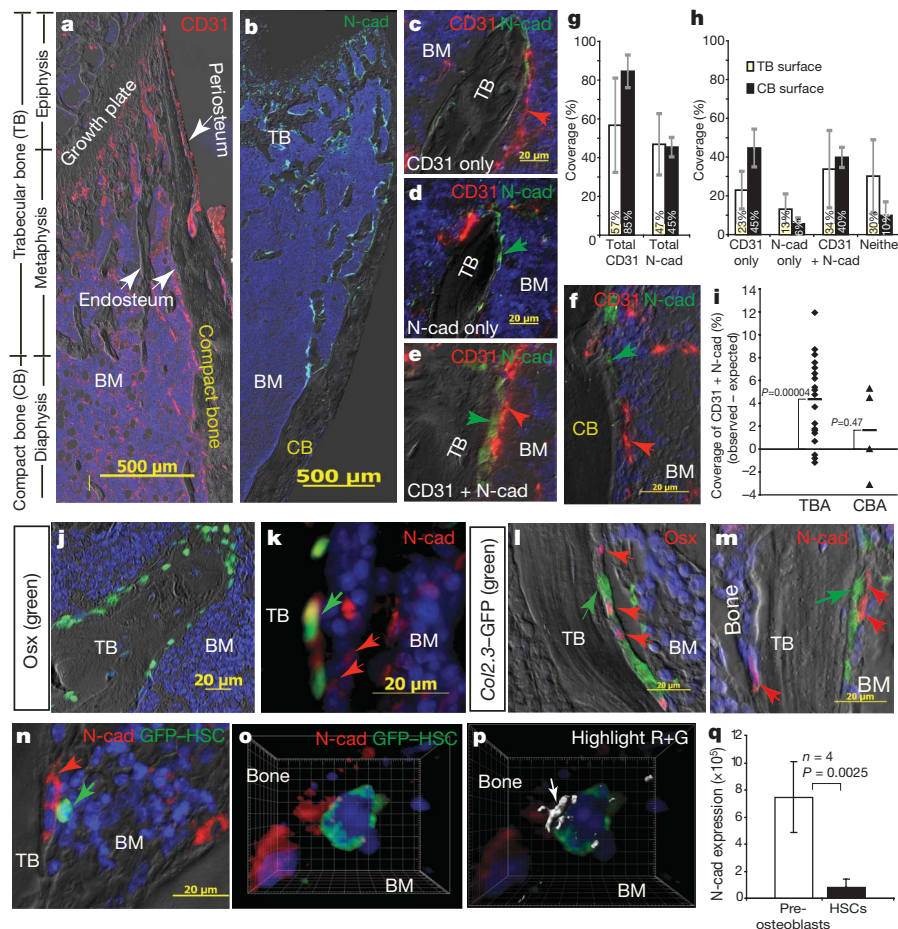
physically distant from the vascular niche<sup>6,19,20</sup>. Therefore, we examined the relationship between osteoblastic and vascular structures and found that vasculature was distributed throughout bone marrow, extending into most endosteal zones (Fig. 2a).

We next determined to what extent N-cadherin<sup>+</sup> osteoblastic cells, a key niche component<sup>2,5</sup>, and vasculature were juxtaposed (Fig. 2b–f, Supplementary Fig. 3 and Supplementary Table 3). The overall distribution of N-cadherin<sup>+</sup> osteoblastic cells in the femur was revealed by immunoassaying (Fig. 2b) and in N-cadherin-LacZ<sup>21</sup> transgenic mice (Supplementary Fig. 4). Overall, the coverage rates by vasculature were higher (57% of trabecular bone area; 85% of compact bone area) than that of N-cadherin<sup>+</sup> osteoblastic cells (47% of trabecular bone area; 45% of compact bone area) (Fig. 2g). This might result in different microenvironments: N-cadherin<sup>+</sup> osteoblastic cells devoid of vasculature (13% of trabecular bone area; 6% of compact bone area); vasculature and N-cadherin<sup>+</sup> cells (23% of trabecular bone area; 45% of compact bone area); and vasculature plus N-cadherin<sup>+</sup> osteoblastic cells (34% of trabecular bone area; 40% of compact bone area)

(Fig. 2h). We calculated the proportion of the endosteum expected to be covered by both vasculature and N-cadherin<sup>+</sup> cells (given the individual frequencies of these properties and assuming independent distribution). We noted that the proportion was higher than expected in trabecular bone area (34% versus 29%,  $P=4.0 \times 10^{-5}$ ) but similar to expectation in compact bone area (40% versus 38%,  $P=0.47$ ) (Fig. 2i and Supplementary Table 4), indicating their non-random co-association in the trabecular bone area.

Osteoblasts have been viewed as a key component of the HSC niche; however, there are different types: pre-osteoblastic cells (stroma-derived (Supplementary Movie 4) and normally inactive but which can be activated to give rise to osteoblasts) and osteoblasts (precursors for osteocytes, a component of the bone structure)<sup>22</sup>. We examined cells expressing a known osteoblast marker, osterix<sup>23</sup> (*Osx*, a nuclear factor, also called *Sp7*), and N-cadherin. *Osx*<sup>+</sup> cells revealed a typical osteoblast distribution pattern (Fig. 2j). Sections co-stained with these two markers showed expression of N-cadherin in some *Osx*<sup>+</sup> cells and in some *Osx*<sup>−</sup> cells that localized either along the bone





surface or adjacent to *Osx*<sup>+</sup> osteoblasts (Fig. 2k). This distribution pattern of N-cadherin<sup>+</sup> cells adjacent to *Osx*<sup>+</sup> cells suggested that N-cadherin<sup>+</sup> cells were more probably pre-osteoblastic cells. We also validated the relationship between N-cadherin<sup>+</sup> cells and osteoblasts (Fig. 2l, m) through staining sections derived from a *Col2.3-GFP*<sup>+</sup> transgenic mouse in which GFP was driven by the collagen-*1* $\alpha$ 2.3 promoter and specifically expressed in osteoblasts<sup>22</sup> (Fig. 2l).

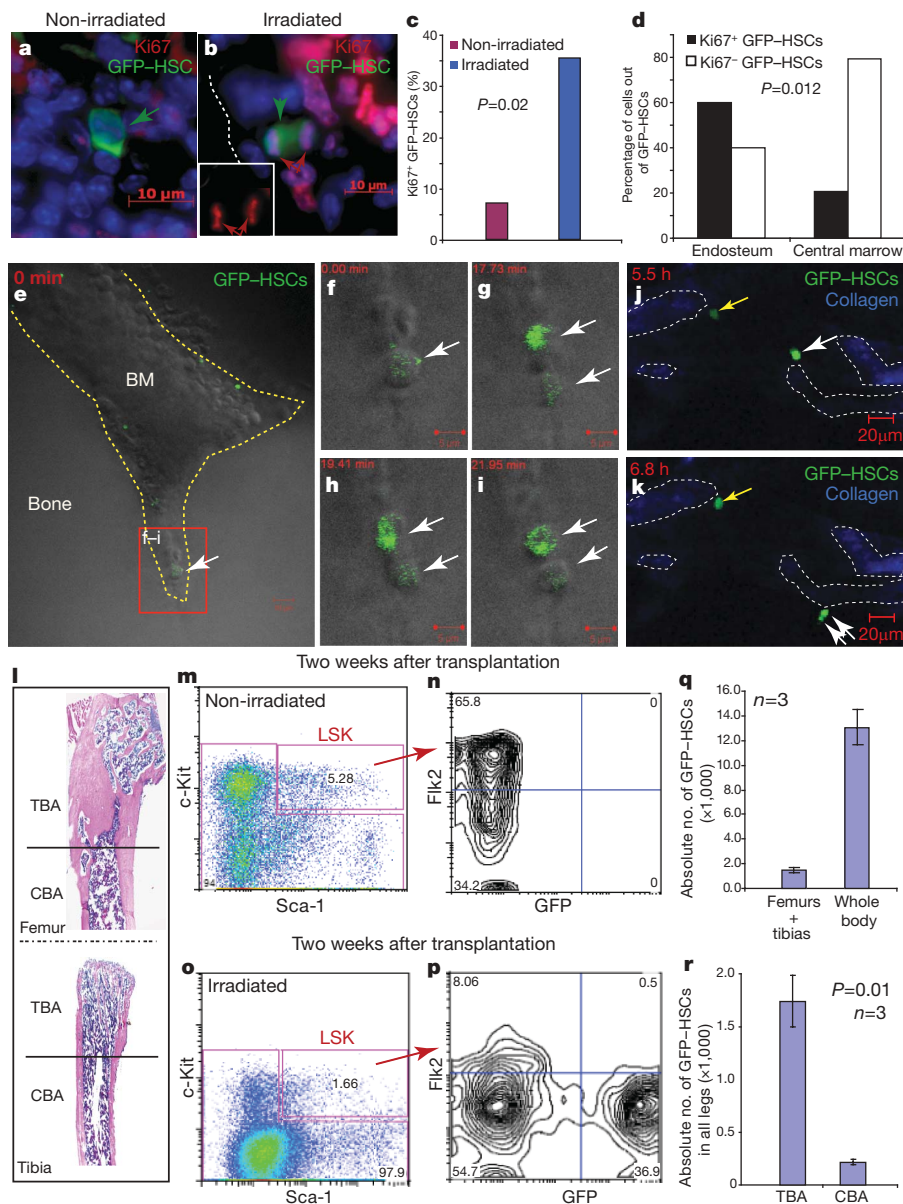
The endosteum offers homing or resident HSCs a range of possible microenvironments as discussed above. To determine with which type of cells GFP<sup>+</sup> HSCs tended to interact, we co-stained with N-cadherin and GFP. GFP<sup>+</sup> HSCs were observed directly attaching to N-cadherin<sup>+</sup> cells (Fig. 2n, o, and Supplementary Fig. 5a) and were close or adjacent to such cells in 10 out of 17 cases (59%; Supplementary Fig. 5a, b). This is consistent with our previous observation of endogenous HSCs (identified using CD201 (ref. 24)) interacting with N-cadherin<sup>+</sup> osteoblastic cells<sup>2,20</sup>. Notably, some of the GFP<sup>+</sup> HSCs were also adjacent to N-cadherin<sup>+</sup> cells in central marrow (Supplementary Fig. 5c). However, not all GFP<sup>+</sup> HSCs were close to N-cadherin<sup>+</sup> cells (Supplementary Fig. 5b), indicating the existence of additional N-cadherin<sup>−</sup> niche components.

Although the level of N-cadherin measured as fluorescence intensity in the GFP<sup>+</sup> HSCs was much lower (average ninefold) compared to adjacent osteoblastic cells (Fig. 2p, q and Supplementary Table 5), its asymmetric distribution at the interface between HSCs and pre-osteoblastic cells (Fig. 2p and Supplementary Movie 5) validates our previous observation<sup>2</sup>.

HSCs normally maintain a steady number but this can be increased in response to bone marrow damage<sup>25</sup>. We compared proliferation of homed GFP<sup>+</sup> HSCs 4–6 h after transplantation using an immunoassay with a marker Ki67 (Fig. 3a–d). In the non-irradiated bone marrow, clusters of proliferating cells were typically seen in the

central marrow, and isolated proliferating cells were often seen in the endosteum (Supplementary Fig. 6a). In the irradiated host, however, proliferating cells were substantially increased in the endosteal region (Supplementary Fig. 6b). Furthermore, we found that GFP<sup>+</sup> HSCs were rarely Ki67<sup>+</sup> in non-irradiated hosts but were five times more often Ki67<sup>+</sup> in irradiated hosts (7% compared with 35%) (Fig. 3a–c and Supplementary Table 6). Notably, Ki67<sup>+</sup>GFP<sup>+</sup> HSCs were frequently found in locations adjacent to clusters of proliferating cells, and this was particularly obvious under irradiated conditions (Fig. 3b). Irradiation-induced changes in the location of proliferating cells were also reflected by more Ki67<sup>+</sup>GFP<sup>+</sup> HSCs but fewer Ki67<sup>−</sup>GFP<sup>+</sup> HSCs being detected in the endosteal compared to the central marrow regions (Fig. 3d and Supplementary Table 7). All these observations indicate that irradiation induced a dynamic change in the endosteal microenvironment from normal inhibition to stress-induced stimulation of proliferation, favouring the expansion of homed GFP<sup>+</sup> HSCs.

Using EVISC, we were also able to observe directly HSC proliferation. We performed this experiment at different time points. At 5–8 h after transplantation (Fig. 3e), a GFP<sup>+</sup> HSC located at the deepest part of a recess in bone gave rise to two daughter cells (Fig. 3f–i and Supplementary Movie 6). In another case, we observed two homed HSCs in the same field of view with different behaviours over 15-h real-time imaging: one was stable and the other was actively mobile and dividing (Fig. 3j, k and Supplementary Movie 8). At 36–40 h after transplantation (Supplementary Movie 7), three out of four GFP<sup>+</sup> HSCs in the same area were stable at the endosteal region. We recorded active division of homed GFP<sup>+</sup> HSCs in three out of twenty-one cases (14.2%) in three independent experiments. Our immunoassays identified the endosteal zone of irradiated recipients as a site of increased HSC proliferation after transplantation. Our



**Figure 3 | Real-time imaging and immunoassaying of homed  $\text{GFP}^+ \text{HSCs}$  and measurement of HSC proliferation and expansion.** **a, b**, Co-staining  $\text{GFP}^+ \text{HSCs}$  with  $\text{Ki67}$ . **c, d**, Percentage (**c**) and distribution (**d**) of proliferating  $\text{GFP}^+ \text{HSCs}$ . (See Supplementary Tables 6 and 7 for statistics.) **e**, Real-time imaging of homed  $\text{GFP}^+ \text{HSCs}$ . Bone structure (DIC). **f–i**, Division of  $\text{GFP}^+ \text{HSCs}$ . **j, k**, Inactive (yellow arrow) and active (white arrow)  $\text{GFP}^+ \text{HSCs}$  during 15-h real-time imaging. Blue indicates bone (dotted line). **l**, Haematoxylin and eosin staining shows separation of trabecular bone area (TBA) and compact bone area (CBA). **m–p**, Measuring the percentage of  $\text{GFP}^+ \text{HSCs}$  ( $\text{Flk2}^+ \text{LSK}$ ). **q**, Absolute number of  $\text{GFP}^+ \text{HSC}$  cells. Calculation of HSC number in whole body based on femurs and tibias containing 15% of total HSCs<sup>6</sup>. **r**, Number of  $\text{GFP}^+ \text{HSCs}$  in trabecular bone area and compact bone area ( $P = 0.01$ , Supplementary Table 8). Error bars indicate s.d.

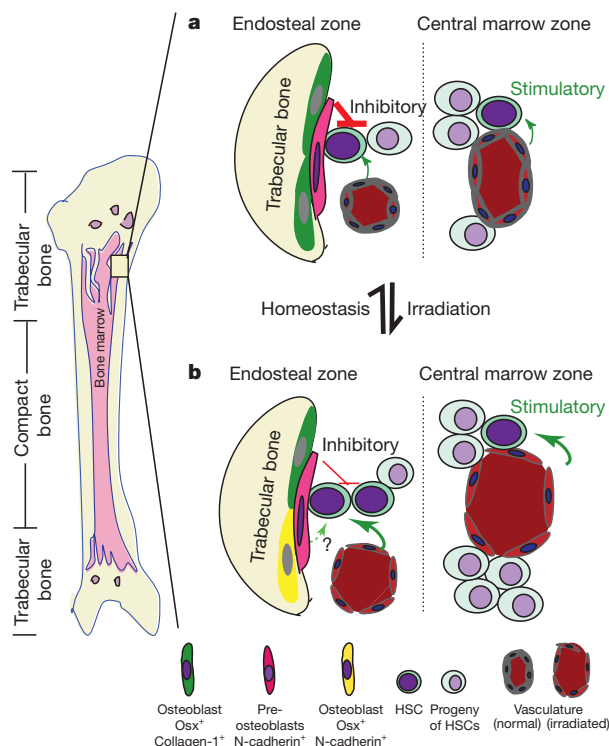
ability to detect division of homed HSCs by EVISC correlates well with this finding. The key advantage of EVISC is that we can monitor the diverse dynamic behaviour of individual HSCs over many hours of real-time imaging versus a snap-shot of a static picture.

To confirm the observed expansion of homed HSCs in irradiated mice, we measured the HSC number using flow cytometry assay 2 weeks after transplantation of 2,000  $\text{GFP}^+ \text{HSCs}$  ( $\text{Flk2}^+ \text{LSK}$ ; Fig. 3l–r). First we found that HSCs were not stable in non-irradiated bone marrow, as  $\text{GFP}^+ \text{HSCs}$  were rarely detected (Fig. 3m, n). In contrast, HSCs transplanted into irradiated mice underwent rapid expansion (Fig. 3o, p). There were 1,959  $\text{GFP}^+ \text{Flk2}^+ \text{LSK}$  cells in all the legs (two femurs plus tibias, 15% of the whole body<sup>6</sup>) determined by flow assay (Fig. 3q), and this equates to 13,065 (1,959, 15%)  $\text{GFP}^+ \text{Flk2}^+ \text{LSK}$  cells in the whole mouse body (Supplementary Table 8). Interestingly, we also found a higher number of expanded HSCs in the trabecular bone area than in the compact bone area (Fig. 3r), separated as indicated in Fig. 3l.

Here we developed an *ex vivo* system to image the homing and proliferation of HSCs in real time, facilitating detection of functional niches and monitoring the behaviour of stem cells. We showed that the endosteal region forms a special zone to maintain HSCs normally but can support their expansion in response to bone marrow damage. We clarified that HSCs tend to (but do not always)

home to and directly interact with N-cadherin<sup>+</sup> pre-osteoblastic cells (Fig. 4a). Osteoblastic-cell-generated signals normally inhibit cell proliferation<sup>26</sup>; however, vascular signals can also be a major component of the endosteal zone (Fig. 4a). Bone marrow damage results in a changed microenvironment that favours HSC transient expansion. We speculate that this might result from an enhanced vascular signal and potentially attenuated inhibitory signals from osteoblastic cells (Fig. 4b). Moreover, signals from osteoblastic cells can directly promote HSC proliferation in the context of mobilization<sup>27</sup> (Fig. 4b). In contrast, the central marrow may form a different zone where dominant vascular signals favour cell proliferation, as has been proposed<sup>19</sup> and is further supported by the transient proliferation of HSCs in the vasculature-dominant spleen in response to bone marrow damage (Figs 1j and 4b). Although we focused on investigating osteoblastic and vascular components, other cell types (osteoclasts<sup>28</sup>, reticular cells<sup>29</sup>, neuronal cells<sup>30</sup>, haematopoietic progenitors) within the immediate environment can also influence the behaviour and states of HSCs. The EVISC method allows us to study the behaviour of individual HSCs and their dynamic interactions with niches in an *ex vivo* setting, and to study the stem cell fate determination in combination with different types of animal models in which HSC marking and lineage tracing are available.





**Figure 4 | Illustration of endosteal zone and central marrow zone.** **a**, The endosteal zone is the inner bone surface and is under homeostasis. HSCs residing in this zone normally receive inhibitory signals from osteoblastic cells but are also exposed to vascular signals. In the central marrow zone, vascular signals might be dominant. **b**, When bone marrow, including HSC, is damaged by irradiation, the endosteal environment transiently converts into a stimulatory environment. This might be due to both reduction in osteoblastic inhibitory and increase in vascular stimulatory signals.

## METHODS SUMMARY

Mice were handled according to Institutional and NIH guidelines. *Ex vivo* imaging was achieved using a confocal laser scanning microscope LSM 510 META equipped with a pulsed NIR laser Chameleon Ultra for two-photon excitation. Fluorescence intensity measurement was processed using the Imaris 6.0.0 software (Bitplane) to facilitate three-dimensional rendering and fluorescence quantification. For immunofluorescent assay, anti-N-cadherin (1:50, YS, IBL), anti-CD31 (1:100, BD Pharmingen), anti-GFP (1:200, Novus Biologicals), anti-Osx (1:200, Abcam) and anti-Ki67 (1:25, Dako) were used. X-gal and TRAP<sup>+</sup> staining followed our previous procedure<sup>20</sup>. To measure the coverage percentage by CD31 vasculature and N-cadherin<sup>+</sup> osteoblasts, for each image, a line was drawn along the bone surface; then we ran a custom java script which encompassed the area within 16  $\mu\text{m}$  of the bone surface, generating a straightened image and creating a line profile from that straightened image. Flow cytometry analysis of haematopoietic cells was described previously<sup>20</sup>. To compare HSCs in trabecular bone area and compact bone area, femurs and tibias were separated before collecting bone marrow. The Scl-TVA mice were injected intravenously with 150 mg kg<sup>-1</sup> of 5-fluorouracil 4 days before injection of RCAS-Luc<sup>18</sup> virus. Bioluminescence imaging of live mice was done using a Xenogen IVIS-200 imager. The live imaging could only detect strong bioluminescence signals, which might be attenuated by thick bone, muscle and fur. For real-time PCR we followed our reported procedure<sup>20</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 28 April; accepted 14 November 2008.  
Published online 3 December 2008.

- Schofield, R. The relationship between the spleen colony-forming cell and the haematopoietic stem cell. *Blood Cells* 4, 7–25 (1978).
- Zhang, J. *et al.* Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* 425, 836–841 (2003).
- Calvi, L. M. *et al.* Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature* 425, 841–846 (2003).
- Nilsson, S. K., Johnston, H. M. & Coverdale, J. A. Spatial localization of transplanted hematopoietic stem cells: inferences for the localization of stem cell niches. *Blood* 97, 2293–2299 (2001).

- Arai, F. *et al.* Tie2/angiopoietin-1 signaling regulates hematopoietic stem cell quiescence in the bone marrow niche. *Cell* 118, 149–161 (2004).
- Kiel, M. J., Yilmaz, O. H., Iwashita, T., Terhorst, C. & Morrison, S. J. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* 121, 1109–1121 (2005).
- Murphy, G. J. *et al.* Manipulation of mouse hematopoietic progenitors by specific retroviral infection. *J. Biol. Chem.* 278, 43556–43563 (2003).
- Schroeder, T. Imaging stem-cell-driven regeneration in mammals. *Nature* 453, 345–351 (2008).
- Christensen, J. L. & Weissman, I. L. Flk-2 is a marker in hematopoietic stem cell differentiation: a simple method to isolate long-term stem cells. *Proc. Natl Acad. Sci. USA* 98, 14541–14546 (2001).
- Juopperi, T. A. *et al.* Isolation of bone marrow-derived stem cells using density-gradient separation. *Exp. Hematol.* 35, 335–341 (2007).
- Denk, W., Strickler, J. H. & Webb, W. W. Two-photon laser scanning fluorescence microscopy. *Science* 248, 73–76 (1990).
- König, K. Multiphoton microscopy in life sciences. *J. Microsc.* 200, 83–104 (2000).
- Zhang, J. *et al.* PTEN maintains haematopoietic stem cells and acts in lineage choice and leukaemia prevention. *Nature* 441, 518–522 (2006).
- Adams, G. B. *et al.* Stem cell engraftment at the endosteal niche is specified by the calcium-sensing receptor. *Nature* 439, 599–603 (2006).
- Jung, Y. *et al.* Regulation of SDF-1 (CXCL12) production by osteoblasts; a possible mechanism for stem cell homing. *Bone* 38, 497–508 (2006).
- Gothert, J. R. *et al.* *In vivo* fate-tracing studies using the Scl stem cell enhancer: embryonic hematopoietic stem cells significantly contribute to adult hematopoiesis. *Blood* 105, 2724–2732 (2005).
- Chen, C. Z. *et al.* Identification of endoglin as a functional marker that defines long-term repopulating hematopoietic stem cells. *Proc. Natl Acad. Sci. USA* 99, 15468–15473 (2002).
- Sausville, J. *et al.* RCAS/SCL-TVA animal model allows targeted delivery of polyoma middle T oncogene to vascular endothelial progenitors *in vivo* and results in hemangioma development. *Clin. Cancer Res.* 14, 3948–3955 (2008).
- Kopp, H. G., Avezilla, S. T., Hooper, A. T. & Rafii, S. The bone marrow vascular niche: home of HSC differentiation and mobilization. *Physiology* 20, 349–356 (2005).
- Haug, J. S. *et al.* N-cadherin expression level distinguishes reserved versus primed states of hematopoietic stem cells. *Cell Stem Cell* 2, 367–379 (2008).
- Luo, Y., Kostetskii, I. & Radice, G. L. N-cadherin is not essential for limb mesenchymal chondrogenesis. *Dev. Dyn.* 232, 336–344 (2005).
- Kalajic, I. *et al.* Use of type I collagen green fluorescent protein transgenes to identify subpopulations of cells at different stages of the osteoblast lineage. *J. Bone Miner. Res.* 17, 15–25 (2002).
- Nakashima, K. *et al.* The novel zinc finger-containing transcription factor osterix is required for osteoblast differentiation and bone formation. *Cell* 108, 17–29 (2002).
- Balazs, A. B., Fabian, A. J., Esmen, C. T. & Mulligan, R. C. Endothelial protein C receptor (CD201) explicitly identifies hematopoietic stem cells in murine bone marrow. *Blood* 107, 2317–2321 (2006).
- Osawa, M. *et al.* *In vivo* self-renewal of c-Kit<sup>+</sup> Sca-1<sup>+</sup> Lin<sup>(low/-)</sup> hemopoietic stem cells. *J. Immunol.* 156, 3207–3214 (1996).
- Nilsson, S. K. *et al.* Osteopontin, a key component of the hematopoietic stem cell niche and regulator of primitive hematopoietic progenitor cells. *Blood* 106, 1232–1239 (2005).
- Mayack, S. R. & Wagers, A. J. Osteolineage niche cells initiate hematopoietic stem cell mobilization. *Blood* 112, 519–531 (2008).
- Kollet, O. *et al.* Osteoclasts degrade endosteal components and promote mobilization of hematopoietic progenitor cells. *Nature Med.* 12, 657–664 (2006).
- Sugiyama, T., Kohara, H., Noda, M. & Nagasawa, T. Maintenance of the hematopoietic stem cell pool by CXCL12-CXCR4 chemokine signaling in bone marrow stromal cell niches. *Immunity* 25, 977–988 (2006).
- Katayama, Y. *et al.* Signals from the sympathetic nervous system regulate hematopoietic stem cell egress from bone marrow. *Cell* 124, 407–421 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We appreciate scientific support from R. Krumlauf, W. Neaves and encouragement from G. Lu and W. Shen. We are grateful to D. Scadden and P. Kulesa for discussion. We thank D. Rowe for providing the Col2.3-GFP line. We thank R. Yu, L. Ma, Q. Qiu and W. Wang for technological advice, M. Hembree, A. Box, H. Marshall, E. Rendenbaugh, C. Cooper and M. Smith for technical support, S. Beckham for histology assistance, J. Perry and J. Ross for comments, and K. Tannen for proofreading. R.A.F. was supported by DOD grant W81XWH-04-1-0801 and by a DRIF Award from the University of Maryland. L.L. is supported by Stowers Institute for Medical Research.

**Author Contributions** Y.X. and T.Y. developed the initial idea, designed and performed experiments. W.W. contributed to EVISC instrumentation and real-time imaging. X.C.H. helped on data analysis, troubleshooting and immunostaining for Ncad LacZ. D.S. assisted in real-time imaging. J.P., J.H., J.W., T.J. and S.Z. assisted in some experiments. J.S., K.P. and R.A. performed quantification of fluorescent intensity. H.L. contributed to statistics. D.M. and R.A.F. contributed to bioluminescent imaging. J.G. assisted in data analysis and manuscript writing. L.L. contributed to overall supervision, experimental design and manuscript writing.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to R.A.F. (rfeldman@umaryland.edu) and L.L. (lil@stowers-institute.org).

## METHODS

**Ex vivo imaging of stem cells.** We isolated the femurs and tibias from the transplanted mice and cut the bone in the vertical orientation (2–3 mm in height, Fig. 1a). The isolated bone (especially trabecular bone) was embedded vertically in 0.5% agarose in DMEM medium (phenol-red free) in a 3.5-cm dish facing the coverslip bottom, allowing the laser beam to penetrate into the bone marrow from the open end of bone/bone marrow (Fig. 1a). All live cell images were acquired using a confocal laser scanning microscope, LSM 510 META (Carl Zeiss), equipped with a pulsed NIR laser Chameleon Ultra (Coherent) for two-photon excitation. Before imaging live cells, we established 920 nm to be the optimal wavelength for two-photon excitation of GFP. We used a 32-channel META detector for spectral analysis and positive identification of GFP-labelled cells. After positive identification, we imaged the cells using an internal PMT detector with a band-pass filter of 500–550 nm, well separating the specific fluorescence signal from autofluorescence and second harmonic signals. The pinhole was set to its most open position. We used the Carl Zeiss objectives Plan Apochromat with a magnification of  $\times 20$  or C-Achroplan NIR with a magnification of  $\times 40$  with a numerical aperture of 0.8. These objectives are optimized for two-photon NIR excitation and a coverslip thickness of 0.170 mm. The images shown in Figs 1 and 3 were not processed except for cropping,  $3 \times 3$  median filter, and linear contrast adjustment. Caution needed to be exercised when a particular area was constantly exposed to the laser beam. The targeted cells might be damaged, resulting in the appearance of some large cells ( $\sim 10$ – $20 \mu\text{m}$ ) converted from GFP<sup>−</sup> into GFP<sup>+</sup> cells by actively fusing with GFP<sup>+</sup> donor cells, indicating that they were probably macrophages (not shown). The 3.5-cm glass bottom culture dishes were from MatTek Corporation.

**Probability to detect HSCs in single femur.** An average adult mouse weights 20 g. Assuming an average density of the mouse body of 1 g per  $1 \text{ cm}^3$  (equal to the density of water) we estimate the volume of a mouse body to be  $20 \times 10^3 \text{ mm}^3$ . A femur is about 10 mm long and has a radius of 0.5 mm, resulting in a volume of  $8 \text{ mm}^3$ . Thus, the volume of a femur is  $4 \times 10^{-4}$  times the volume of a mouse body. If we inject 50,000 HSCs and assume even distribution throughout the whole mouse body, we would expect to find 20 cells per femur.

**Immunohistochemical and immunofluorescent assay.** The femurs and tibias were fixed in Zn<sup>2+</sup>-formalin and processed for paraffin sections. For immunofluorescent staining, after antigen retrieval using EZ Retriever Microwave (BioGenex), nonspecific antibody binding was blocked by incubating slides with  $1 \times$  universal blocking reagent (BioGenex, HK085-5k) at room temperature for 1 h. Anti-N-cadherin (1:50, YS, IBL), anti-CD31 (1:100 dilution, BD Pharmingen), anti-GFP (1:200, Novus Biologicals), anti-Osx (1:200, Abcam), or anti-Ki67 (1:25, Dako) was added to the tissue section slides. YS polyclonal antibody was used for immunoassaying because the monoclonal antibody

MNCD2 was not efficient. However, we cannot exclude the possibility that YS antibody may recognize other cadherins in stromal cells, as MNCD2 is not efficient to block the staining of the non-LSK cells out of lineage-negative bone marrow. The slides were incubated at 4 °C overnight. Donkey anti-rabbit 546/568/488 or anti-rat 546/568 or donkey anti-goat 488 Alexa Fluor conjugate (1:200, Invitrogen) was added at room temperature for 1 h. For CD31 staining, we used the TSA Biotin System (PerkinElmer, NEL700). DAPI was used to stain the nucleus and the slides were then observed under a fluorescent microscope. X-gal and TRAP<sup>+</sup> staining has been described previously<sup>2,20</sup>.

**Fluorescence intensity measurement.** The data sets were processed using Imaris 6.0.0 software (Bitplane) to facilitate three-dimensional rendering and fluorescence quantification. The Surpass Scene in Imaris was used to create a contour surface object that encapsulated cells of interest based on DAPI fluorescence signal. The mean fluorescence intensity of the channel of interest in each contour surface was recorded. A manual background subtraction was then performed by measuring the mean red intensity of a uniform background area and subtracting the background mean intensity value from each of the cell intensity mean values.

**Analysis of CD31 and N-cadherin coverage of the endosteum.** Expected 'double positive' coverage of the endosteum by both CD31 and N-cadherin, assuming they are independently distributed, was calculated for each line profile as the product  $p \times q$  where  $p$  and  $q$  are the total fractions of the line profile covered by CD31 and N-cadherin, respectively. The observed and expected double positive rates were compared using two-tailed paired  $t$ -tests.

**Flow cytometry analysis of HSCs.** Bone marrow cells were flushed from the femurs and tibias and the red blood cells were lysed using a 0.15 M ammonium chloride solution. HSCs were enriched using FACS as lineage PeCy5-negative (the lineage cocktail contained CD3, CD4, CD5, CD8, Gr1, CD11b, B220, IgM and Ter119), Sca-1 PE-positive, c-Kit APC-positive, Biotin-Flk2 streptavidin PEcy7-negative. For transplantation experiments, donor HSCs were isolated from GFP transgenic mice (Jackson lab). Rescue bone marrow cells ( $2 \times 10^5$ ) were from Ptpcr.

To compare HSCs in trabecular bone area and compact bone area, two ends of femurs and one end of tibias were cut as trabecular bone area (2–3 mm away from the end); the other end of tibia (3–4 mm away from the end) was cut and discarded; and the middle part of femurs and tibias was combined as compact bone area. Bone marrow cells were flushed from trabecular bone area and compact bone area separately then followed the regular HSCs analysis protocol.

**Real-time RT-PCR.** We followed the procedure reported previously<sup>24</sup>. Primers used for amplification of *Sdf1* were as follows: forward, 5'-GGACGCCAAGGT-CGTCGCCGT-3'; reverse, 5'-TAATTCGGGTCAATGCACA-3'. HPRT was used as an internal control for normalization of input mRNA.



## LETTERS

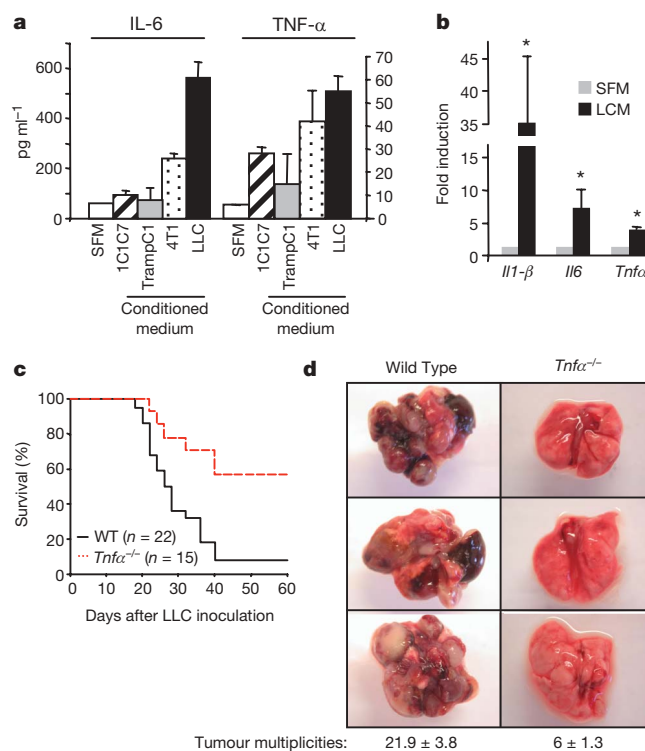
# Carcinoma-produced factors activate myeloid cells through TLR2 to stimulate metastasis

Sunhwa Kim<sup>1</sup>, Hiroyuki Takahashi<sup>1</sup>, Wan-Wan Lin<sup>1,2</sup>, Pascal Descargues<sup>1</sup>, Sergei Grivennikov<sup>1</sup>, Youngjun Kim<sup>1†</sup>, Jun-Li Luo<sup>1†</sup> & Michael Karin<sup>1</sup>

Metastatic progression depends on genetic alterations intrinsic to cancer cells as well as the inflammatory microenvironment of advanced tumours<sup>1,2</sup>. To understand how cancer cells affect the inflammatory microenvironment, we conducted a biochemical screen for macrophage-activating factors secreted by metastatic carcinomas. Here we show that, among the cell lines screened, Lewis lung carcinoma (LLC)<sup>3</sup> were the most potent macrophage activators leading to production of interleukin-6 (IL-6) and tumour-necrosis factor- $\alpha$  (TNF- $\alpha$ ) through activation of the Toll-like receptor (TLR) family members<sup>4</sup> TLR2 and TLR6. Both TNF- $\alpha$  and TLR2 were found to be required for LLC metastasis. Biochemical purification of LLC-conditioned medium (LCM) led to identification of the extracellular matrix proteoglycan versican, which is upregulated in many human tumours including lung cancer<sup>5,6</sup>, as a macrophage activator that acts through TLR2 and its co-receptors TLR6 and CD14. By activating TLR2:TLR6 complexes and inducing TNF- $\alpha$  secretion by myeloid cells, versican strongly enhances LLC metastatic growth. These results explain how advanced cancer cells usurp components of the host innate immune system, including bone-marrow-derived myeloid progenitors<sup>7</sup>, to generate an inflammatory microenvironment hospitable for metastatic growth.

Distant-site metastases are the leading cause of cancer-associated mortality. They depend on genetic and/or epigenetic alterations that are intrinsic to cancer cells, or extrinsic factors provided by the tumour microenvironment<sup>1</sup>. For instance, cytokines produced by inflammatory cells can enhance metastatogenesis by repressing the metastasis suppressor maspin within primary prostate carcinoma cells<sup>8</sup>. Furthermore, tumour progression and metastasis positively correlate with presence of infiltrates containing myeloid and lymphoid cells<sup>2,9</sup>. It has been shown that certain carcinomas secrete factors that upregulate fibronectin and recruit vascular endothelial growth factor receptor 1 (VEGFR1)-positive haematopoietic progenitors to sites of future metastatic growth, termed the pre-metastatic niche<sup>7</sup>. To examine whether cancer cells secrete factors that directly activate myeloid cells to produce tumour-promoting cytokines<sup>10</sup>, we collected serum-free conditioned medium from different cancer cell lines, derived mainly from C57BL/6 mice, and applied it to bone-marrow-derived macrophages (BMDM), which were assayed for production of interleukin-1 $\beta$  (IL-1 $\beta$ ), IL-6 and TNF- $\alpha$ . The screen included 1C1C7 and TrampC1, which are liver and prostate cancer cell lines, respectively, with little or no metastatic activity, and two metastatic breast and lung carcinomas, 4T1 and LLC, respectively. Conditioned medium from metastatic cells, especially LLC, induced higher amounts of IL-6 and TNF- $\alpha$  secretion than conditioned medium from non-metastatic cells (Fig. 1a). IL-1 $\beta$  secretion was

undetectable and the conditioned medium did not contain IL-6 or TNF- $\alpha$  (data not shown). LCM also induced expression of *Il1 $\beta$* , *Il6* and *Tnfa* messenger RNAs (mRNAs), whereas serum-free medium (SFM) and NIH3T3-conditioned medium were inactive (Fig. 1b and data not shown). We investigated the metastatogenic function of some of the LCM-induced cytokines by tail vein injection of LLC into age- and sex-matched *Tnfx* and *Il6* knockout mice and wild-type (WT) controls. *Tnfx*<sup>-/-</sup> mice exhibited markedly ( $P < 0.001$ ) reduced



**Figure 1 | Metastatic carcinomas secrete factors that induce macrophage production of TNF- $\alpha$ , needed for lung metastasis.** **a**, BMDM were cultured with SFM or SFM conditioned by mouse carcinoma cells (conditioned medium), and cytokine production was measured (averages  $\pm$  s.d.,  $n = 3$ ). **b**, BMDM were cultured with SFM or LCM, and cytokine mRNAs were measured by quantitative PCR with reverse transcription. Fold-induction above SFM-treated cells was determined (averages  $\pm$  s.d.,  $n = 3$ ; \*,  $P < 0.001$  by Student's  $t$  test). **c**, Survival of WT ( $n = 22$ ) and *Tnfx*<sup>-/-</sup> ( $n = 15$ ) mice inoculated with LLC ( $1 \times 10^6$  cells) through the tail vein ( $P < 0.001$ ; log-rank test for significance). **d**, Lungs of WT and *Tnfx*<sup>-/-</sup> mice 47 days after LLC inoculation ( $2 \times 10^5$  cells). Tumour multiplicities are shown underneath (averages  $\pm$  s.d.,  $n = 11$ ,  $P < 0.001$  by Student's  $t$ -test).

<sup>1</sup>Department of Pharmacology and Cancer Center, School of Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0723, USA. <sup>2</sup>Department of Pharmacology, College of Medicine, National Taiwan University, Taipei, China. <sup>†</sup>Present addresses: Department of Applied Biochemistry, Konkuk University, 322 Danwol-dong, Chungju-City, Chungbuk 380-701, Korea (Y.K.); Department of Cancer Biology, The Scripps Research Institute, 5353 Parkside Drive, Jupiter, Florida 33458, USA (J.-L.L.).

mortality compared with WT mice after inoculation with  $1 \times 10^6$  LLC cells (Fig. 1c) and showed an even greater survival advantage when given a smaller LLC inoculum (Supplementary Fig. 1a). Similar differences were seen in lung tumour multiplicity (Fig. 1d). By contrast, there was little difference in survival of *Il6*<sup>-/-</sup> and WT mice inoculated with  $1 \times 10^6$  LLC cells (Supplementary Fig. 1b). Thus, TNF- $\alpha$  but not IL-6 is important for LLC metastasis.

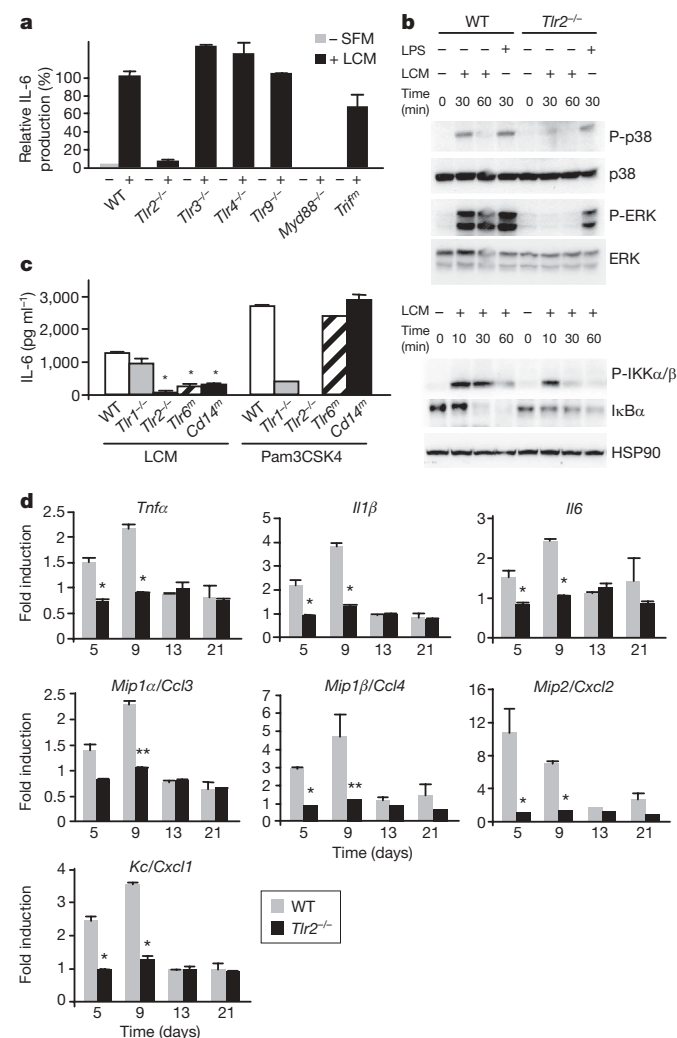
We explored the involvement of TLR family members in sensing LCM components. BMDM from mice deficient in TLR2, TLR3, TLR4 or TLR9 or their adaptor proteins, Myd88 and TRIF (which is inactivated by the *Lps2* mutation: *Trif*<sup>tm</sup>)<sup>4</sup>, were examined for production of IL-6, a convenient BMDM activation marker. LCM-induced IL-6 was fully dependent on TLR2 and Myd88 but not on TLR3, TLR4, TLR9 or TRIF (Fig. 2a). *Tlr2*<sup>-/-</sup> BMDM were also defective in LCM-induced *Il1 $\beta$*  and *Il6* mRNA expression and LCM did not induce anti-tumorigenic type I interferon genes

(Supplementary Fig. 2a), which are readily induced after TLR3 or TLR4 engagement<sup>11</sup>. TLR2 was also required for LCM-induced IL-6 and TNF- $\alpha$  secretion by alveolar macrophages, which produced tenfold more TNF- $\alpha$  than BMDMs (Supplementary Fig. 2b). TLR2 was required for optimal LCM-induced activation of mitogen-activated protein kinases and I $\kappa$ B kinase or I $\kappa$ B $\alpha$  degradation (Fig. 2b). TLR2 uses TLR1, TLR6 or CD14 as co-receptors<sup>12</sup>. LCM-induced IL-6 production was dependent on TLR2, TLR6 and CD14 but not on TLR1 (Fig. 2c). By contrast, the response to Pam3CSK4, a bacterial lipoprotein analogue<sup>4</sup>, depended on TLR2 and TLR1 but not on TLR6 and CD14. These results rule out possible contamination with bacterial lipoproteins. Furthermore, anti-mycoplasma treatment of LLC had no effect on LCM activity (data not shown).

To examine the *in vivo* role of TLR2, we inoculated sex- and age-matched *Tlr2*<sup>-/-</sup> and WT mice with  $2 \times 10^5$  LLC cells through the tail vein and measured mRNAs encoding cytokines and chemokines in their lungs. LLC-induced lung inflammation has been previously described<sup>13</sup>, but its mechanism was unknown. *Tnfx*, *Il1 $\beta$* , *Il6* and inflammatory chemokine mRNAs were induced 5 days after LLC inoculation in WT lungs, peaking at 9 days after inoculation (Fig. 2d). None of these mRNAs was induced in lungs of *Tlr2*<sup>-/-</sup> mice, whose basal content of *Mip1 $\beta$ /Ccl4*, *Mip2/Cxcl2* and *Kc/Cxcl1* mRNAs was higher than WT lungs (Supplementary Fig. 3). In addition, *Tlr2*<sup>-/-</sup> and WT mice were subcutaneously inoculated with LLC cells and examined for lung macrophage infiltration and inflammatory cytokine gene expression days later. Although no difference was observed in primary subcutaneous tumour growth, macrophage infiltration and inflammatory cytokine gene expression were greatly reduced in *Tlr2*<sup>-/-</sup> mice compared with WT mice (Supplementary Fig. 4a–c).

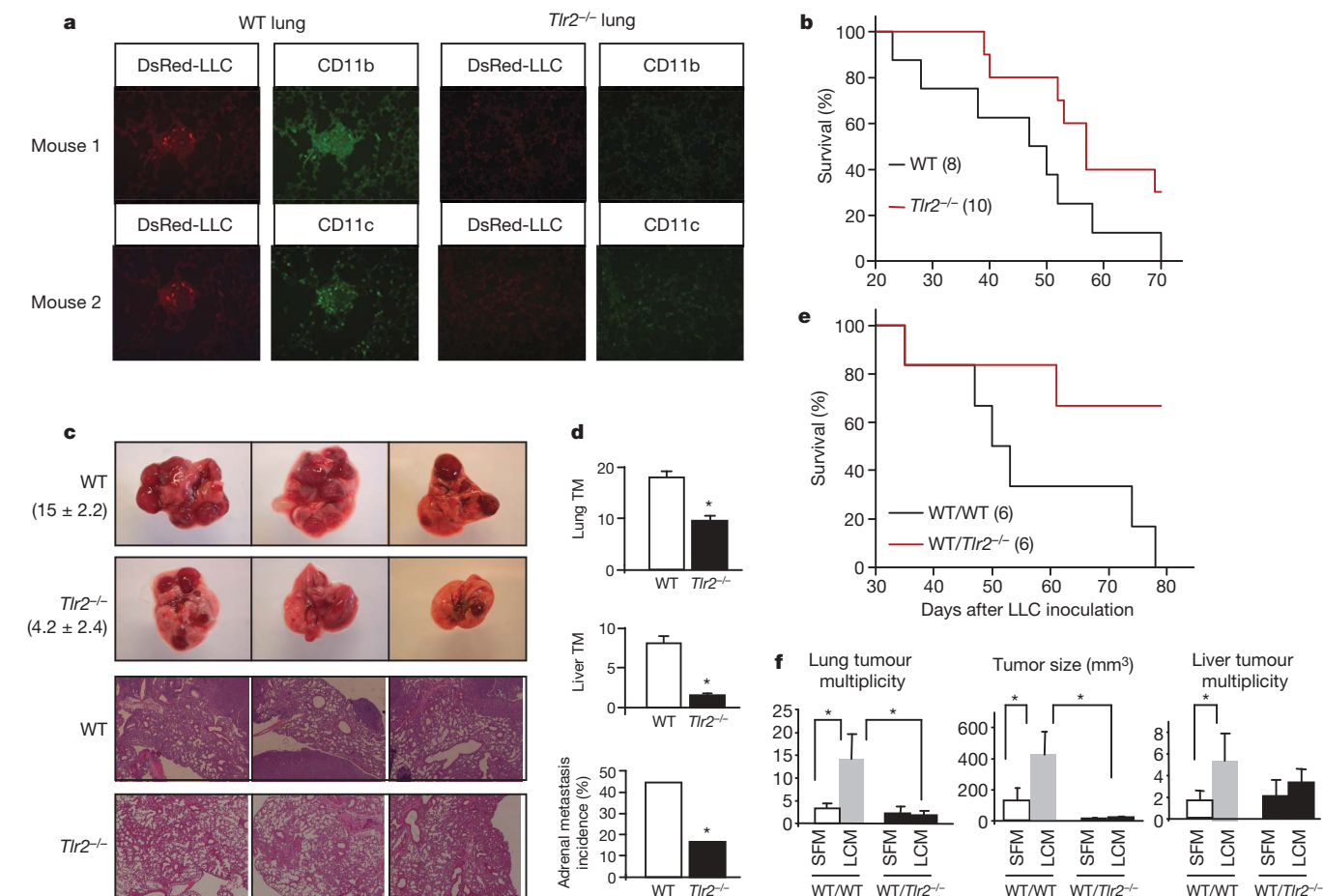
To investigate whether TLR2 signalling contributes to LLC metastatogenesis, we inoculated age- and sex-matched *Tlr2*<sup>-/-</sup> and WT mice with dsRed-labelled LLC cells through the tail vein and examined their lungs for micrometastases. WT but not *Tlr2*<sup>-/-</sup> lungs showed small clusters of DsRed-LLC cells with adjacent CD11b<sup>+</sup> and CD11c<sup>+</sup> myeloid cells (Fig. 3a). We also detected a few cells that carried the CD3 antigen (T cells) in micro-metastases of WT lungs (data not shown). Importantly, *Tlr2*<sup>-/-</sup> mice exhibited significantly greater ( $P < 0.02$ ) survival than WT mice after LLC inoculation and their lungs contained fewer and smaller tumour nodules (Fig. 3b, c). Tumour nodules in WT mice contained more CD11b<sup>+</sup>/Gr1<sup>+</sup> inflammatory monocytes/myeloid suppressors and IL-10<sup>high</sup>/F4/80<sup>+</sup> M2 macrophages (Supplementary Fig. 5). *Tlr2*<sup>-/-</sup> mice exhibited significantly fewer lung and liver tumour nodules than WT mice and a lower incidence of adrenal gland metastasis after subcutaneous implantation of LLC cells (Fig. 3d). To investigate whether TLR2 acts in bone-marrow-derived cells, we examined survival of LLC-inoculated chimaeric mice. Mice reconstituted with *Tlr2*<sup>-/-</sup> bone marrow (WT/*Tlr2*<sup>-/-</sup>) exhibited markedly improved ( $P < 0.04$ ) survival compared with mice reconstituted with WT bone marrow (WT/WT) (Fig. 3e). WT/WT and WT/*Tlr2*<sup>-/-</sup> mice were also inoculated with  $2 \times 10^5$  LLC cells followed by intraperitoneal injections of LCM or SFM. Lung and liver tumour loads were significantly higher ( $P < 0.05$ ) in WT/WT mice receiving LCM than in those receiving SFM along with the LLC inoculum (Fig. 3f). The pro-metastatic effect of LCM was dependent on TLR2 activation, as little or no metastatic enhancement was seen in WT/*Tlr2*<sup>-/-</sup> mice. These results strongly suggest that LCM contains TLR2-activating factors that enhance metastatogenesis.

To identify the nature of these factors, we collected large amounts of LCM, separated it on a mono-Q anion exchange column and monitored column fractions for their ability to induce IL-6 in BMDM. Fractions with IL-6-inducing activity were pooled and separated on a Superdex 200 sizing column (Supplementary Fig. 6). Most of the IL-6-inducing activity eluted in a few high molecular mass (greater than 400 kDa) fractions that contained several polypeptides larger than 200 kDa. These fractions were pooled, deglycosylated and



**Figure 2 | LLC-secreted factor activates TLR2 to induce lung inflammation.** **a**, BMDM from indicated mouse strains were cultured with SFM or LCM, and IL-6 production was measured (*m*: mutant allele; averages  $\pm$  s.d.,  $n = 3$ , presented as percentage of WT LCM-stimulated value). **b**, BMDM were treated with LCM or LPS (100 ng ml<sup>-1</sup>). Cell lysates were examined for kinase phosphorylation (P) and I $\kappa$ B $\alpha$  degradation by immunoblotting. Total ERK and HSP90 are loading controls. **c**, BMDM from indicated mouse strains were cultured with LCM or Pam3CSK4 (1 ng ml<sup>-1</sup>), and IL-6 production was measured (averages  $\pm$  s.d.,  $n = 4$ ). **d**, RNA was extracted from lungs of WT or *Tlr2*<sup>-/-</sup> mice at indicated times after LLC inoculation ( $2 \times 10^5$  cells). mRNAs were quantified as above and the amounts in non-inoculated WT or *Tlr2*<sup>-/-</sup> lungs were given a value of 1.0 (averages  $\pm$  s.e.m.,  $n = 3$ ; \* $P < 0.05$ , \*\* $P < 0.005$  (compared with WT) by Student's *t*-test).





**Figure 3 | TLR2 is required for metastatic growth.** **a**, WT and *Tlr2*<sup>-/-</sup> lungs were analysed 9 days after inoculation of DsRed-LLC ( $2 \times 10^5$  cells) for DsRed and myeloid cells markers (CD11b, CD11c) using fluorescence microscopy (magnification  $\times 200$ ). **b**, Survival of LLC inoculated ( $2 \times 10^5$  cells) WT ( $n = 8$ ) and *Tlr2*<sup>-/-</sup> ( $n = 10$ ) mice ( $P < 0.02$ ; log-rank test for significance). **c**, Lungs and haematoxylin-and-eosin-stained lung sections (magnification  $\times 25$ ) 20 days after LLC inoculation. Tumour multiplicity is shown on the left (averages  $\pm$  s.e.m.,  $n = 8$ ,  $P < 0.001$  by Student's *t*-test). **d**, Tumour multiplicities of lung and liver metastatic nodules and incidence

of adrenal metastasis 17 days after primary tumour removal (averages  $\pm$  s.e.m., WT  $n = 9$ , *Tlr2*<sup>-/-</sup>  $n = 6$ ;  $P < 0.05$  by Student's *t*-test). **e**, Survival of WT/WT or WT/*Tlr2*<sup>-/-</sup> chimaeric mice inoculated with LLC ( $2 \times 10^5$  cells) 6–7 weeks after bone-marrow reconstitution ( $P < 0.04$ ; log-rank test for significance;  $n = 6$ ). **f**, Lung tumour multiplicity (left), size (middle) and liver tumour multiplicity (right) in chimaeric mice, 27 (lung) or 48 (liver) days after LLC injection ( $2 \times 10^5$  cells) together with SFM or LCM (averages  $\pm$  s.e.m.,  $n = 8$ ;  $P < 0.05$  by Student's *t*-test).

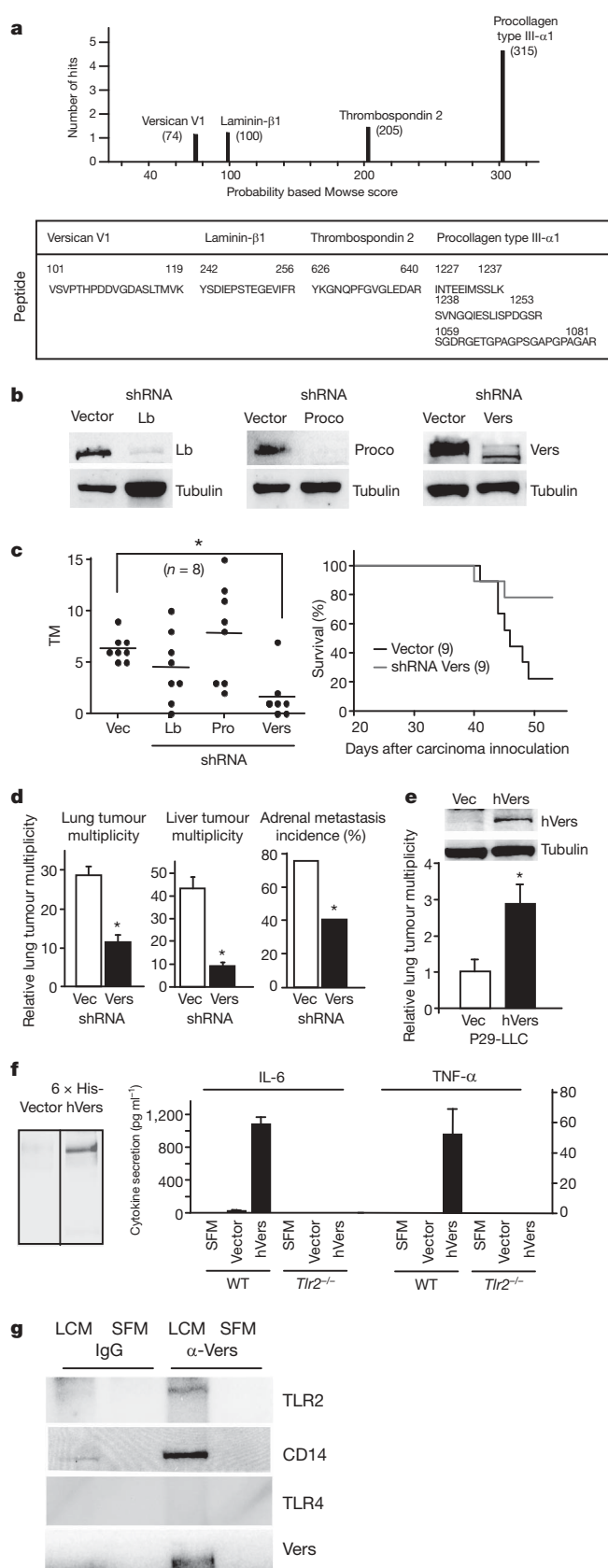
subjected to mass spectrometry, resulting in identification of several peptides derived from the extracellular matrix proteins: versican V1, laminin- $\beta$ 1, thrombospondin 2 and procollagen type III- $\alpha$ 1 (Fig. 4a). To examine which protein accounts for induction of inflammatory cytokines, we incubated LCM with individual neutralizing antibodies before BMDM stimulation and measurement of cytokine production. Incubation of LCM with antibodies to versican, laminin- $\beta$ 1 or procollagen type III- $\alpha$ 1, but not thrombospondin 2, reduced IL-6 and TNF- $\alpha$  production (Supplementary Fig. 7). A control antibody to HMGB1, an inflammatory mediator released by necrotic cells<sup>14</sup>, did not inhibit cytokine induction.

To investigate the role of these proteins in LCM-enhanced metastatogenesis, we generated stable LLC cell lines containing short hairpin RNA (shRNAs) specific to versican V1, laminin- $\beta$ 1 or procollagen III- $\alpha$ 1. Silencing efficiency was approximately 90% or higher (Fig. 4b). The silenced cells and LLC cells transduced with a control shRNA were injected into mice and lung tumour nodules and their survival monitored. Silencing of versican V1 significantly reduced ( $P < 0.001$ ) tumour multiplicity, but silencing of laminin- $\beta$ 1 resulted only in a modest inhibition, whereas procollagen III- $\alpha$ 1 silencing slightly enhanced tumour multiplicity (Fig. 4c). Furthermore, silencing of versican V1 reduced lung nodule multiplicity fourfold (Supplementary Fig. 8a). Tumours isolated from mice inoculated with versican-silenced LLC cells displayed very low versican expression

(Supplementary Fig. 8b), whereas lung tumours from LLC-inoculated mice expressed more versican than normal lung (Supplementary Fig. 8c). Importantly, mice inoculated with versican-silenced LLC cells exhibited significantly improved ( $P < 0.05$ ) survival (Fig. 4c). Silencing of versican also reduced metastatic spread to lung, liver and adrenal glands in the subcutaneous implantation model (Fig. 4d). To ascertain the proinflammatory and premetastatic functions of versican, we used the low metastatic LLC variant P29-LLC<sup>15</sup>. Conditioned medium from P29-LLC did not induce IL-6 in BMDM and contained very little versican (Supplementary Fig. 9). Ectopic expression of human versican in P29-LLC cells increased lung tumour multiplicity after tail vein injection (Fig. 4e).

To examine how versican activates macrophages, we produced His-tagged human versican V1 in LLC cells and purified it on an Ni-chelate column. The purified protein induced IL-6 and TNF- $\alpha$  production in WT but not *Tlr2*<sup>-/-</sup> BMDM (Fig. 4f). We investigated whether versican interacts with TLR2 or its CD14 co-receptor. Immunoprecipitation of lysates of LCM-incubated Raw264.7 macrophages with versican-specific antibody, but not a control antibody, co-precipitated TLR2 and CD14 but not TLR4 (Fig. 4g). The versican antibody did not precipitate TLR2 unless the macrophages were first incubated with LCM.

Metastasis is the result of a complex process involving invasion of adjacent tissues, intravasation, circulatory transport, arrest at a



distant site, extravasation, growth, survival and neoangiogenesis<sup>16</sup>. Bone-marrow-derived cells, such as macrophages<sup>2</sup> and haematopoietic progenitors<sup>7</sup>, are important participants in this process; however, how they are mobilized and activated to support metastasis is unclear. Our results indicate that versican secretion by LLC cells is necessary for metastatic spread to lung, liver and adrenal gland, a process that

#### Figure 4 | Versican is a TLR2 agonist and metastasis-enhancing factor.

**a**, Identification of candidate LCM macrophage activators by mass spectrometry. Probability-based Mowse scores (upper) and tryptic peptides corresponding to the indicated proteins (lower). **b**, LLC were transduced with indicated shRNA lentiviruses. After selection, expression of the indicated proteins was analysed. **c**, Left: LLC transduced with indicated lentiviruses were injected ( $2 \times 10^5$  cells) into mice ( $n = 8$ ). Lung tumour multiplicity was enumerated at 20 days (averages  $\pm$  s.e.m.,  $P < 0.001$  by Student's *t*-test). Right: survival of mice inoculated with LLC transduced with control or Vers shRNAs ( $n = 9$ ;  $*P < 0.05$ ; log-rank test for significance). **d**, Lung and liver tumour multiplicities and incidence of adrenal metastasis 17 days after primary tumour removal (averages  $\pm$  s.e.m.,  $n = 8$  and  $n = 5$  for mice injected with control or Vers shRNA-transduced cells, respectively;  $*P < 0.05$  by Student's *t*-test). **e**, Upper: non-transfected and human versican (hVers) transfected P29-LLC cells were analysed for versican expression. Lower: hVers non-expressing and expressing P29-LLC cells were injected through the tail vein ( $n = 7$ ). After 27 days, lung metastatic nodules were enumerated ( $*P < 0.05$  by Student's *t*-test). **f**, Left: silver staining of purified  $6 \times$  His-hVers. Right: BMDM were incubated without or with His-hVers for 20–24 h, and cytokine secretion was measured (averages  $\pm$  s.d.,  $n = 3$ ). **g**, Lysates of Raw264.7 cells incubated with LCM or SFM for 1 h were immunoprecipitated with versican-specific or control antibody and analysed by immunoblotting with the indicated antibodies.

depends on TLR2-mediated myeloid cell activation and TNF- $\alpha$  production. Versican is an aggregating chondroitin sulphate proteoglycan that accumulates both in tumour stroma and cancer cells<sup>5,6</sup>. It can bind hyaluronan, and both versican and hyaluronan are highly expressed in non-small-cell lung cancer (NSCLC), especially in advanced disease with high recurrence rate, whereas versican in normal lung is rather low<sup>6</sup>. Versican or fragments thereof enhance tumour cell migration, growth and angiogenesis, processes that are of direct relevance to metastasis<sup>17</sup>. Versican also binds to several adhesion molecules expressed by inflammatory cells and has pro-inflammatory activity<sup>18</sup>. A related extracellular matrix proteoglycan, biglycan, has been reported to activate both TLR2 and TLR4<sup>19</sup>, but our results indicate that the pro-inflammatory activities of versican rely on TLR2 but not TLR4. TLR2 recognizes Gram-positive bacteria-derived lipoteichoic acid and lipoproteins<sup>4</sup>. This activity mainly depends on TLR2:TLR1 dimers<sup>20</sup>, but the response to versican requires TLR6 and not TLR1 as a co-receptor. Although TLR2 and versican interact, it is not clear whether this interaction is direct or depends on a versican ligand, such as hyaluronan. Indeed, hyaluronan fragments can activate macrophages through TLR2 (ref. 21), and hyaluronan accumulation and the enzyme that converts large hyaluronan polymers to smaller fragments, hyaluronidase, have been linked to metastasis<sup>22</sup>.

TLR2 on host myeloid cells and their product TNF- $\alpha$  are important positive modulators of LLC metastatic behaviour, although neither protein influences primary tumour growth of subcutaneously implanted LLC. It appears that TNF- $\alpha$  is one of the major pro-metastatic factors produced by host myeloid cells. TNF- $\alpha$  can suppress the apoptosis of cancer cells and stimulate their proliferation through NF- $\kappa$ B activation<sup>23</sup>. In addition, by increasing vascular permeability<sup>24</sup>, TNF- $\alpha$  can enhance recruitment of leukocytes as well as intravasation and extravasation of cancer cells. We suggest that versican, its interaction with TLR2 and production of TNF- $\alpha$  by activated myeloid cells provide potential points for anti-metastatic intervention.

#### METHODS SUMMARY

Detailed methods are given in Supplementary Information. Briefly, LLC cells were injected through the tail vein or subcutaneously implanted into 6- to 7-week-old mice at  $2 \times 10^5$  to  $2 \times 10^6$  cells per mouse to measure metastases to lung, liver or adrenal gland. Metastasis-enhancing factors were purified from LCM by column chromatography, and identified on a QSTAR XL qTOF mass spectrometer. Factor activity was determined by the ability to induce IL-6 production by BMDM. Gene and protein expression were monitored by quantitative PCR and immunoblot analysis, respectively. Tumours and their composition were analysed by immunohistochemistry and indirect immunofluorescence.



**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 1 August; accepted 20 October 2008.**

1. Kopstein, L. & Christofori, G. Metastasis: cell-autonomous mechanisms versus contributions by the tumor microenvironment. *Cell. Mol. Life Sci.* **63**, 449–468 (2006).
2. Pollard, J. W. Tumour-educated macrophages promote tumour progression and metastasis. *Nature Rev. Cancer* **4**, 71–78 (2004).
3. Weiss, L. & Ward, P. M. Lymphogenous and hematogenous metastasis of Lewis lung carcinoma in the mouse. *Int. J. Cancer* **40**, 570–574 (1987).
4. Akira, S., Uematsu, S. & Takeuchi, O. Pathogen recognition and innate immunity. *Cell* **124**, 783–801 (2006).
5. Isogai, Z. *et al.* 2B1 antigen characteristically expressed on extracellular matrices of human malignant tumors is a large chondroitin sulfate proteoglycan, PG-M/versican. *Cancer Res.* **56**, 3902–3908 (1996).
6. Pirinen, R. *et al.* Versican in nonsmall cell lung cancer: relation to hyaluronan, clinicopathologic factors, and prognosis. *Hum. Pathol.* **36**, 44–50 (2005).
7. Kaplan, R. N. *et al.* VEGFR1-positive haematopoietic bone marrow progenitors initiate the pre-metastatic niche. *Nature* **438**, 820–827 (2005).
8. Luo, J. L. *et al.* Nuclear cytokine-activated IKK $\alpha$  controls prostate cancer metastasis by repressing Maspin. *Nature* **446**, 690–694 (2007).
9. DeNardo, D. G., Johansson, M. & Coussens, L. M. Immune cells as mediators of solid tumor metastasis. *Cancer Metastasis Rev.* **27**, 11–18 (2008).
10. Lin, W. W. & Karin, M. A cytokine-mediated link between innate immunity, inflammation, and cancer. *J. Clin. Invest.* **117**, 1175–1183 (2007).
11. Hsu, L. C. *et al.* The protein kinase PKR is required for macrophage apoptosis after activation of Toll-like receptor 4. *Nature* **428**, 341–345 (2004).
12. Bas, S. *et al.* The proinflammatory cytokine response to *Chlamydia trachomatis* elementary bodies in human macrophages is partly mediated by a lipoprotein, the macrophage infectivity potentiator, through TLR2/TLR1/TLR6 and CD14. *J. Immunol.* **180**, 1158–1168 (2008).
13. Hiratsuka, S. *et al.* MMP9 induction by vascular endothelial growth factor receptor-1 is involved in lung-specific metastasis. *Cancer Cell* **2**, 289–300 (2002).
14. Scaffidi, P., Misteli, T. & Bianchi, M. E. Release of chromatin protein HMGB1 by necrotic cells triggers inflammation. *Nature* **418**, 191–195 (2002).
15. Nakanishi, H. *et al.* Structural differences between heparan sulphates of proteoglycan involved in the formation of basement membranes *in vivo* by Lewis-lung-carcinoma-derived cloned cells with different metastatic potentials. *Biochem. J.* **288**, 215–224 (1992).
16. Fidler, I. J. The pathogenesis of cancer metastasis: the ‘seed and soil’ hypothesis revisited. *Nature Rev. Cancer* **3**, 453–458 (2003).
17. Zheng, P. S. *et al.* Versican/PD-M G3 domain promotes tumor growth and angiogenesis. *FASEB J.* **18**, 754–756 (2004).
18. Wight, T. N. Versican: a versatile extracellular matrix proteoglycan in cell biology. *Curr. Opin. Cell Biol.* **14**, 617–623 (2002).
19. Schaefer, L. *et al.* The matrix component biglycan is proinflammatory and signals through Toll-like receptors 4 and 2 in macrophages. *J. Clin. Invest.* **115**, 2223–2233 (2005).
20. Buwitt-Beckmann, U. *et al.* Toll-like receptor 6-independent signaling by diacylated lipopeptides. *Eur. J. Immunol.* **35**, 282–289 (2005).
21. Scheibner, K. A. *et al.* Hyaluronan fragments act as an endogenous danger signal by engaging TLR2. *J. Immunol.* **177**, 1272–1281 (2006).
22. Lokeshwar, V. B., Cerwinka, W. H., Itoyama, T. & Lokeshwar, B. L. HYAL1 hyaluronidase in prostate cancer: a tumor promoter and suppressor. *Cancer Res.* **65**, 7782–7789 (2005).
23. Luo, J. L., Maeda, S., Hsu, L. C., Yagita, H. & Karin, M. Inhibition of NF-kappaB in cancer cells converts inflammation- induced tumor growth mediated by TNF $\alpha$  to TRAIL-mediated tumor regression. *Cancer Cell* **6**, 297–305 (2004).
24. Tracey, K. J. *et al.* Shock and tissue injury induced by recombinant human cachectin. *Science* **234**, 470–474 (1986).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** S.K. was supported by the International Human Frontier Science Program Organization (IHFSPO), the National Cancer Institute-sponsored Cancer Therapeutic Training Program and a Ruth L. Kirschstein National Research Service Award. Y.K., H.T., J.-L.L., P.D. and S.G. were supported by the California Institute of Regenerative Medicine, the Japanese Respiratory Society, the Life Science Research Foundation, IHFSPO and the Crohn's and Colitis Foundation of America, respectively. Work in the M.K. laboratory was supported by grants from the National Institutes of Health and a Littlefield-AACR grant in Metastatic Colon Cancer Research. M.K. is an American Cancer Society Research Professor. We thank S. Akira and B. Beutler for TLR and adaptor protein deficient mice, D. Zimmermann for human versican construct, R. Hoffman for DsRed-LLC cells and K. Takenaga for LLC-P29 cells. We also thank R. Gallo and K. Yamasaki for the hyaluronan-inhibiting peptides and advice, J. Varner for analysing cell migration and Santa Cruz Biotechnology for gifts of antibodies.

**Author Contributions** S.K., H.T., W.-W.L. and M.K. conceived the project and planned experiments and analyses, which were performed by S.K., H.T. and W.-W.L. Y.K. and J.-L.L. helped with protein purification and tail-vein injection of cancer cells and tumour analysis, respectively. P.D. and S.G. analysed M2 macrophages and tissue versican content, and effect of TNF- $\alpha$  neutralization on lung metastasis. M.K. oversaw the entire project and wrote the manuscript with S.K.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to M.K. ([karinoffice@ucsd.edu](mailto:karinoffice@ucsd.edu)).

## METHODS

**Animal and cell culture.** *Tlr1*<sup>-/-</sup>, *Tlr2*<sup>-/-</sup>, *Tlr3*<sup>-/-</sup>, *Tlr4*<sup>-/-</sup>, *Tlr9*<sup>-/-</sup>, *Myd88*<sup>-/-</sup> and *Trif*<sup>m</sup>, *Tlr6*<sup>m</sup> and *Cd14*<sup>m</sup> (other names: *Lps*, *Insouciant* and *Heedless*, respectively) mutant mice were provided by S. Akira (Osaka University) and B. Beutler (The Scripps Research Institute). *Tnfr*<sup>-/-</sup> and *Il6*<sup>-/-</sup> mice were from the Jackson Laboratory. All strains used in metastasis studies were backcrossed to C57BL6 mice more than ten times. Mice were maintained under specific pathogen-free conditions, and experimental protocols were approved by the University of California, San Diego Animal Care Program, following National Institutes of Health guidelines. LLC, 1C1C7, TrampC1 and 4T1 cells were from the American Type Culture Collection. DsRed-LLC cells were provided by R. Hoffman (Anti-Cancer, Inc.). The cells were cultured under standard conditions, except for collection of conditioned media, and were confirmed to be mycoplasma free.

**Metastasis assays.** Formation of lung and liver metastases by cancer cells inoculated through the tail vein was examined as described<sup>23</sup>. Briefly, subconfluent LLC cells were harvested and passed through a 40-µm cell strainer, washed three times in PBS, re-suspended in PBS and inoculated at 2 × 10<sup>5</sup> or 1 × 10<sup>6</sup> cells per 7- to 8-week-old mouse through the tail vein. At designated time points, mice were killed, and their lungs were removed, weighed and histologically examined. Some mice were kept until death and survival data were obtained. Lung tumour nodules were microdissected using an 18G needle under a microscope for protein and RNA analysis. Alternatively, the entire tumour-bearing lung was used. Subcutaneous cancer cell implantation was as described<sup>25</sup>. Briefly, 2 × 10<sup>6</sup> control or versican-silenced LLC cells were subcutaneously injected into the right flank. When necessary, the primary tumour was surgically removed using aseptic techniques at 3–4 weeks after inoculation, and lung, liver and adrenal gland metastases were analysed 2–3 weeks after removal of the primary tumour.

**Bone marrow transplantation.** C57BL6 mice were lethally irradiated (1050 rad) and transplanted with 1 × 10<sup>6</sup> bone marrow cells from WT or *Tlr2*<sup>-/-</sup> mice.

**DsRed visualization, immunohistochemistry, histological analysis and flow cytometry.** Tissues were immediately frozen in OCT compound (Tissue-Tek) and post-fixed with acetone. Sections were mounted with Vectashield containing DAPI (4,6-diamidino-2-phenylindole), and observed under a fluorescent microscope (Zeiss) to detect DsRed-labelled cells. The sections were also incubated with FITC-labelled antibodies to CD11b, CD11c and CD3 (all from Pharmingen) to detect inflammatory cells. Histological analysis of M2 macrophages was conducted as described<sup>26</sup>. Briefly, tumour-bearing lungs were fixed in 10% neutral buffered formalin. Paraffin-embedded lung sections (5 µm) were kept at 60 °C for 1 h, and then incubated at 94 °C for 40 min in Dako's antigen retrieval solution S2368. Slices were stained with primary antibodies against IL-10 (Santa Cruz Biotechnology) and F4/80 (Caltag) for 1 h at room temperature. A specific signal was detected using the appropriate Dako EnVision System, HRP (DAB) kit. CD11b<sup>+</sup>/Gr1<sup>+</sup> myeloid suppressors or inflammatory monocytes present in lungs were analysed by flow cytometry using CD11b-Alexa 647 and Gr1-PE-Cy7 (eBioscience) antibodies and an ACCURI flow cytometer.

**Conditioned media.** Conditioned medium was collected from cells incubated in SFM for 20 h and filtered through a 0.22-µm filter. Conditioned medium samples were added to BMDM for 16–20 h, after which inflammatory cytokine gene expression or secretion were assayed. Conditioned medium (300 µl) were injected intraperitoneally three times, once every three days into WT mice that had received WT or *Tlr2*<sup>-/-</sup> bone marrow transplants 6–7 weeks before being inoculated with LLC cells. After 27 days, mice were killed and lungs were removed and histologically analysed by staining sections with Harris haematoxylin and eosin.

**Analysis of gene expression.** Total tissue RNA was prepared using RNeasy kit (Qiagen). Quantitative PCR was performed as described<sup>23</sup>. Primer sequences for *Tnfr*, *Il1β*, *Il6* and *Cph* and RT-PCR conditions were as previously described<sup>27</sup>.

Primer pairs for *Mip1α/Ccl3*, *Mip1β/Ccl4*, *Mip2/Cxcl2* and *Kc/Cxcl1* were 5'-TCTGCCGGTTTCTCTTAGTCA-3'/5'-ACCATGACACTCTGCAACCA-3', 5'-CTGCCTCTTTGGTCAGGAA-3'/5'-TTCTGTGCTCCAGGGTTCTC-3', 5'-CTAGCTGCCTGCCTCATTCTAC-3'/5'-CAACAGTGTACYYACGCAGACG-3', 5'-CTTGGGGACACCTTTTAGCA-3'/5'-GCTGGGATTACCTCAA-3'.

Cells and tumour tissues were lysed and analysed by SDS-polyacrylamide gel electrophoresis and immunoblotting<sup>23</sup> with antibodies to p38, IκBα, procollagen III-α1, laminin-β1, tubulin (all from Santa Cruz Biotechnology), ERK, P-ERK, P-p38, P-IKKα/β (all from Cell Signalling, Inc.), versican V1 (Abcam), actin (Sigma) and hsp90 (Pharmingen). IL-6 and TNF-α were measured by enzyme-linked immunosorbent assay (R&D and Pharmingen).

**Lentiviral and retroviral transduction.** siRNAs to mouse versican V1, laminin-β1 and procollagen III-α1 mRNAs were generated as described<sup>8</sup>, cloned into pLSLPw, provided by I. Verma (The Salk Institute), and transfected into 293T cells. siRNA sequences for indicated genes were as follows: versican V1, GTACACAGTTGATGAAATA and TATTTTCATCAACTGTGTAC; laminin-β1, ACAGATACTTCGCCTACGA and TCGTAGGCGCAAGTATCTGT; procollagen III-α1, CCAGAACCATGTCAAATAT and ATATTTGACATGGTCTGG. Human versican V1 vector was provided by D. Zimmermann (University of Zurich) and transfected into GP2-293T cells (Clontech). Lentivirus and retrovirus stocks were prepared as described<sup>8</sup>. Virus-containing supernatants were incubated with LLC cells for 2 days with polybrene, and infected cells were selected in 5 µg ml<sup>-1</sup> puromycin (Invitrogen) for shRNA or 10 µg ml<sup>-1</sup> Zeocin (Invitrogen) for versican expression.

**Protein purification.** Protein purification was conducted using a ÄKTA FPLC (GE Healthcare) system. Initially, samples were loaded onto a Mono-Q HR 5/5 column (Amersham Bioscience) that was developed with a gradient of 0–1 M NaCl in 20 mM Tris-HCl pH 8.0 buffer (flow rate 1 ml min<sup>-1</sup>). Fractions possessing IL-6 inducing activity were collected for further analysis and separation on a HiLoadTM 16/60 SuperdexTM 200 gel filtration column (GE Healthcare) eluted with 20 mM Tris-150 mM NaCl pH 8.0 buffer. Molecular masses were estimated by calibrating the column with a molecular mass marker kit (Sigma). Active fractions were pooled and analysed by SDS-polyacrylamide gel electrophoresis and mass spectrometry.

**Mass spectrometry.** Samples were separated on a 1D NuPage 4–12% Bis-Tris gradient gel (Invitrogen) in MOPS buffer (Invitrogen) and stained with GelCode Blue. Protein bands were excised and in-gel digested with trypsin. The resulting peptides were extracted and separated on an LC Packings liquid chromatographic system equipped with a 75-µm C18 column at a flow rate of 300 nl min<sup>-1</sup>. The column was interfaced through a nanospray source directly into a QSTAR XL QTOF mass spectrometer. A 1-s survey scan was followed by 3-s product ion scans for the two most abundant parent ions. The resulting mass spectrometry data were analysed using MASCOT searching of the Swiss Prot database. The MASCOT search was performed allowing one miscleavage and common modifications of methionine oxidation and amino-terminal pyroglutamic acid.

**Statistical analyses.** Results are expressed as means ± s.e.m. or s.d. Data were analysed by Student's *t*-test and two-way analysis of variance using the GraphPad Prism statistical program. *P* values < 0.05 were considered significant. Error bars depict s.e.m. or s.d.

25. O'Reilly, M. S. et al. Angiostatin: a novel angiogenesis inhibitor that mediates the suppression of metastases by a Lewis lung carcinoma. *Cell* **79**, 315–328 (1994).
26. Urošević, M., Kamarashev, J., Burg, G. & Dummer, R. Primary cutaneous CD8<sup>+</sup> and CD56<sup>+</sup> T-cell lymphomas express HLA-G and killer-cell inhibitory ligand, ILT2. *Blood* **103**, 1796–1798 (2004).
27. Eckmann, L., Fierer, J. & Kagnoff, M. F. Genetically resistant (Ityr) and susceptible (Itys) congenic mouse strains show similar cytokine responses following infection with *Salmonella dublin*. *J. Immunol.* **156**, 2894–2900 (1996).



# Chaperonin complex with a newly folded protein encapsulated in the folding chamber

D. K. Clare<sup>1</sup>, P. J. Bakkes<sup>2†</sup>, H. van Heerikhuizen<sup>2</sup>, S. M. van der Vies<sup>2</sup> & H. R. Saibil<sup>1</sup>

A subset of essential cellular proteins requires the assistance of chaperonins (in *Escherichia coli*, GroEL and GroES), double-ring complexes in which the two rings act alternately to bind, encapsulate and fold a wide range of nascent or stress-denatured proteins<sup>1–5</sup>. This process starts by the trapping of a substrate protein on hydrophobic surfaces in the central cavity of a GroEL ring<sup>6–10</sup>. Then, binding of ATP and co-chaperonin GroES to that ring ejects the non-native protein from its binding sites, through forced unfolding or other major conformational changes, and encloses it in a hydrophilic chamber for folding<sup>11–15</sup>. ATP hydrolysis and subsequent ATP binding to the opposite ring trigger dissociation of the chamber and release of the substrate protein<sup>3</sup>. The bacteriophage T4 requires its own version of GroES, gp31, which forms a taller folding chamber, to fold the major viral capsid protein gp23 (refs 16–20). Polypeptides are known to fold inside the chaperonin complex, but the conformation of an encapsulated protein has not previously been visualized. Here we present structures of gp23–chaperonin complexes, showing both the initial captured state and the final, close-to-native state with gp23 encapsulated in the folding chamber. Although the chamber is expanded, it is still barely large enough to contain the elongated gp23 monomer, explaining why the GroEL–GroES complex is not able to fold gp23 and showing how the chaperonin structure distorts to enclose a large, physiological substrate protein.

Chaperonin–substrate (binary) complexes were formed by rapidly mixing urea-denatured gp23 with GroEL. Ternary complexes were generated by adding gp31 (bacteriophage T4 GroES homologue) and the ATP transition state analogue ADP·AlF<sub>3</sub>. Cryo-electron microscopy (cryo-EM) data sets of 30,000–35,000 particles were collected of both preparations, and initial three-dimensional maps were obtained by treating each data set as a single structure with seven-fold symmetry. The resulting maps showed GroEL and GroEL–gp31 complexes with some additional densities in the binding cavities (Supplementary Fig. 1). As in our earlier study on malate dehydrogenase (MDH) folding, we expected the non-native substrate to form heterogeneous and asymmetric complexes with the chaperonins. Therefore we used a combination of multivariate statistical analysis (MSA) and competitive projection matching to sort the images into more homogeneous classes and determine their structures, without imposing any symmetry<sup>10,21</sup>.

The binary complexes were resolved into five classes (Supplementary Fig. 2, Supplementary Table 1). In agreement with our mass spectrometry results<sup>22,23</sup>, we observed empty GroEL (20% of the images) and GroEL with extra density in either one (40%) or both rings (40%). The classes displaying the largest fraction of the substrate density are shown in Fig. 1, with the GroEL subunit domains fitted into the maps. The GroEL rings deviate little from seven-fold symmetry (Fig. 1c, d, g, h and rotational correlation analysis, not shown).

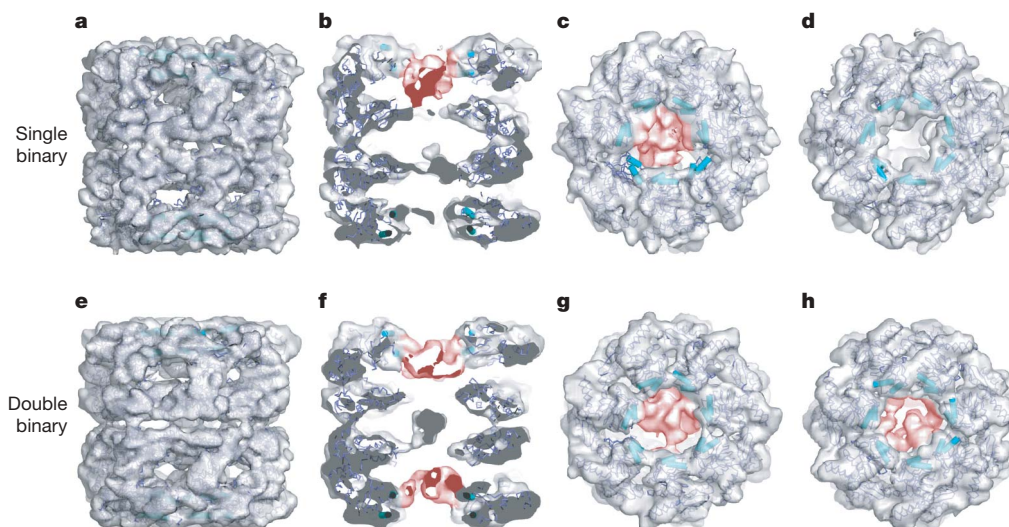
This is unlike GroEL–MDH complexes in which the GroEL apical domains are bunched together on the side of the ring where MDH binds<sup>10</sup>. The large gp23 densities contact at least five of the seven GroEL apical domains in the ring, and the density is located deep inside the cavity, mainly around helix I and the underlying hydrophobic segment<sup>7</sup> (Fig. 1b, c, f–h). Whether substrate is bound or not, the open rings are very similar to one another and to apo GroEL.

Most of the ternary complexes fell into three well-defined structural classes (Fig. 2, Supplementary Table 1): ‘empty’ (no apparent substrate density, 33% of the images), substrate bound in the *trans* (open) ring (32%), and substrate bound in both *cis* and *trans* rings (35%). Consistent with the presence of the ATP analogue, the *trans* ring of all three complexes displayed the same intra-ring  $\beta$ -sheet contact between neighbouring equatorial domains as in the GroEL–GroES–ATP complex<sup>20,24</sup>, although there is some variation in orientation of the apical domains. Less substrate density is resolved in the *trans* ring than in the binary complexes but it is found in a very similar location, that is, deep in the cavity and interacting with most of the apical domains (Fig. 2e, f, h and i). In the *cis* chamber, the substrate density fills most of the space, without strong contacts to either GroEL or gp31 (Fig. 2h). At lower density thresholds, the *cis* substrate density extends to contact several regions of the gp31 lid and the GroEL apical domains, potentially exerting force on the chamber (not shown). The atomic structure of gp23 is not known, but sequence alignment and genetic analysis have shown that it is closely related to the T4 vertex protein, gp24, for which the structure has been determined<sup>25</sup>. The encapsulated gp23 density is remarkably similar in shape to the major domain of gp24 (Fig. 3c).

Unexpectedly, we found that the *cis* chamber of the *trans*-occupied complex is slightly compressed, estimated as a 7% decrease in volume, and is significantly expanded in the *cis/trans* occupied complex, with a 12% increase in volume, relative to the *cis* chamber in the empty complex. This can be seen by the differences in density of the mobile loops of gp31, which interact with the apical domains of GroEL (arrows, Fig. 3a–c), and in the cross-sections through the *cis* apical domains which are almost losing contact at some positions (Fig. 3d–f). The expansion suggests that the folding substrate is exerting pressure on the folding chamber. No density is visible for the small, mobile insertion domain of gp23, but there is space at the base of chamber to accommodate it in a disordered form (Fig. 3c). The compression of the folding chamber in the *trans*-bound complex reveals an allosteric effect of the *trans* substrate, and suggests a mechanism for newly bound substrates to prime dissociation of the folding chamber. The nature of this allosteric effect was surprising, because the connections between GroEL and gp31 are stronger, rather than weaker as expected in preparation for the release of gp31.

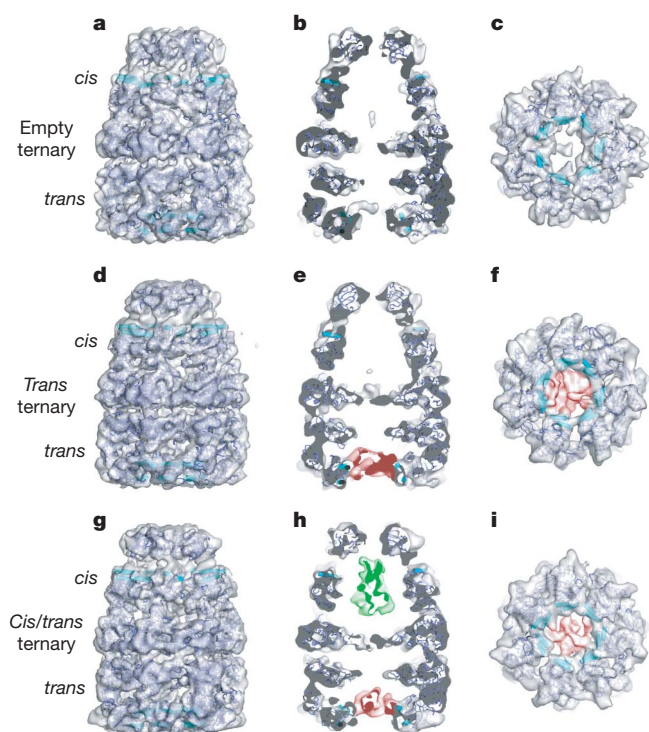
Extracting the gp23 densities from the maps revealed remarkably well-defined structures for both the initially captured, non-native

<sup>1</sup>Department of Crystallography and Institute for Structural and Molecular Biology, Birkbeck College, Malet Street, London WC1E 7HX, UK. <sup>2</sup>Department of Pathology, VU University Medical Center, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands. †Present address: Department of Molecular Microbiology, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Kerklaan 30, 9751 NN Haren, The Netherlands.



**Figure 1 | Asymmetric reconstructions of GroEL with non-native gp23 in one or both rings.** **a–d**, Binary complex with gp23 in one ring (red density), with the crystal structures of GroEL domains fitted into the maps, shown from the side (**a**), as a central section (**b**), from the top (**c**) and from the bottom (**d**). **e–h**, Same views of the binary complex with gp23 in both rings (red density). The EM density maps were sharpened between 20 and 10 Å. Automated docking of the atomic coordinates of the 42 GroEL domains as rigid bodies into each complex gave excellent fits to the maps, with hinge

residues of neighbouring domains in proximity to each other, except for a few regions such as the intermediate domains of some subunits. Helices H and I are shown as cyan cylinders. The carboxy termini of some subunits are visible and either contact the substrate (**b**, upper ring) or bend away from it (**f**, lower ring). Interaction of the flexible GroEL C termini with substrates is consistent with earlier reports<sup>29,30</sup>. The diameter of GroEL complexes is around 140 Å and the resolution of the maps is around 11 Å at 0.5 Fourier shell correlation.



**Figure 2 | Asymmetric reconstructions of GroEL-gp31 without visible substrate, with gp23 in the open ring and with gp23 in both rings.** **a–c**, Empty ternary complex, with the crystal structures of GroEL and gp31 fitted as 49 individual domains, shown from the side (**a**), as a central section (**b**) and from the bottom (**c**). **d–f**, Same views of the *trans*-bound ternary complex. **g–i**, Same views of the *cis/trans* bound ternary complex. Density of gp23 in the *trans* rings is shown in red and in the *cis* ring in green. The resolution of the maps is around 10 Å at 0.5 Fourier shell correlation, and they were therefore sharpened between 20 and 10 Å. The GroEL and gp31 domain coordinates fit very well into the density maps, with only a few minor mismatches.

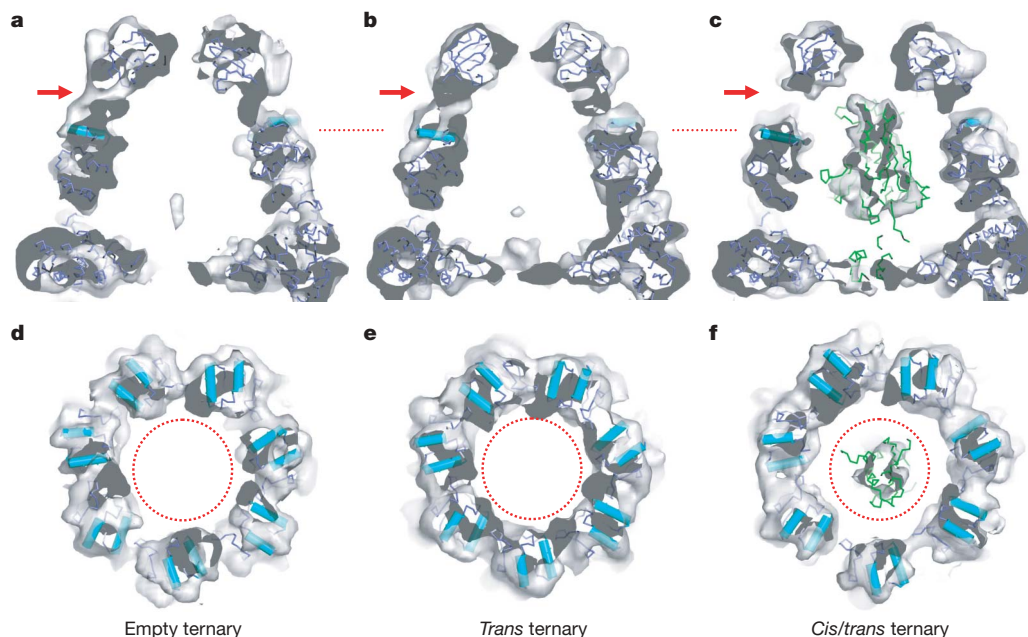
form and the almost fully folded state in the folding chamber of the GroEL–gp31 complex (Fig. 4). The cryo-EM density for gp23 bound to the top rings of the different binary complexes represents over half of the native molecular volume (Fig. 4a, b), showing that the non-native protein is constrained in position and shape when bound to the chaperonin complex (disordered regions are not seen in these averaged structures). The substrate densities in the second ring of doubly occupied complexes (bottom rings) are less well defined, possibly because one substrate dominates the alignment and classification, and the second substrate is not necessarily in a fixed position relative to the first. In the *trans* occupied ternary complex, the observation of less substrate density suggests that it is less well ordered in this class (Fig. 4c). In the *cis* complex, the encapsulated density is clearly recognizable as a low resolution version of the major domain in the gp24 crystal structure (Fig. 4d, e).

A recent FRET (fluorescence resonance energy transfer) study shows that another large, stringent GroEL substrate, Rubisco, occupies a more compact conformation in the *trans* ring of a ternary complex than in the open ring of a binary complex<sup>14</sup>. Our findings do not show a more compact conformation for gp23 in the *trans* ring of the ternary complex. Moreover, our mass spectrometry data shows that Rubisco binds to GroEL with strong negative cooperativity between the rings, but that gp23 binds to both rings<sup>23,26</sup>. Therefore, these two large substrates may have distinct modes of interaction with GroEL. A diversity of GroEL–substrate interaction modes can be anticipated, depending on the folding pathways and intermediates of different substrate proteins.

It is estimated that all of the GroEL in *E. coli* would be required to fold the large number of gp23 subunits produced during T4 infection<sup>27</sup>. It has been proposed that gp23 monopolizes the cell's GroEL through specific amino-terminal regions of gp23 that pause its translation and target the nascent chain to GroEL<sup>27</sup>. Our observation of well-defined electron density for the large domain of gp23 bound to GroEL, in a relatively small number of structural classes, is consistent with the notion that there is a specific binding region on gp23 that targets it to GroEL, leading to a preferred mode of binding.

In our previous study of GroEL–MDH complexes, most of the MDH density was observed in a similar region of the GroEL surface





**Figure 3 | Folding chambers of the GroEL-gp23-gp31 complexes.** **a–c**, Side view sections and **d–f**, cross-sections through the apical domains of the folding chambers (at the position of the red dotted lines in **a–c**). Shown are views of the empty complex (**a** and **d**), the *trans*-occupied complex (**b** and **e**) and the *cis/trans* occupied complex (**c** and **f**). The red arrows in **a–c** show the loss of density at the contact between gp31 and GroEL. Correspondingly, the dotted red circles in **d–f** are all the same size (45 Å in diameter) and

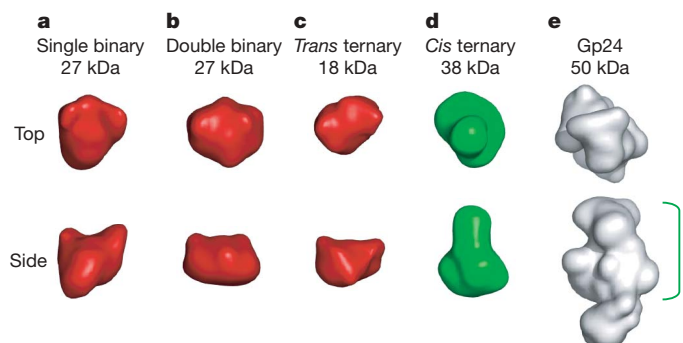
highlight the expansion of the apical domain ring in the *cis/trans* complex and the contraction of the *trans*-occupied apical domain ring. The major domain of the gp24 structure fits very well into the *cis* substrate density (**c**, **f**), unlike the mobile insertion domain of gp24 in the extended conformation seen in the crystal structure, where it makes an inter-subunit contact. In this position it clashes with the GroEL C termini, but there is clearly space available for it closer to the major domain.

to what we report here for gp23, namely the lower part of helix I and the underlying segment<sup>10</sup>. But MDH (33 kDa) only interacts with three of the seven apical domains, whereas gp23 (56 kDa) contacts at least five of the seven sites and occupies more of the binding cavity than MDH. The more asymmetric interaction of MDH with the cavity appears to perturb the GroEL symmetry in both rings<sup>10</sup>. The internal location of the substrate around helix I and the underlying segment leaves the helix H/I groove accessible to the mobile loop of the co-chaperonin even with a large substrate protein bound in the open cavity. However, it was reported that glutamine synthetase (51 kDa) occupies a more external binding site<sup>28</sup>. It is possible that

the external binding site preferentially binds more folded states, similar to that suggested for one class of MDH complexes<sup>10</sup>.

Once gp23 is bound to GroEL, gp31 (but not GroES) can encapsulate and fold the capsid protein<sup>19</sup>. The inability of GroES to encapsulate gp23 can be explained by the elongated shape adopted by newly folded gp23 inside the GroEL-gp31 chamber, and implies that the folding intermediate formed upon ATP and co-chaperonin binding is too bulky or too extended for GroES to bind. In this respect, it should be noted that gp31 has longer mobile loops and a larger internal space, which we show here to be stretched to the limit in order to encapsulate gp23.

The present work gives the first visualization of a newly folded substrate protein trapped in a largely native conformation inside the folding chamber of GroEL. The well-defined gp23 density in the GroEL-gp31 *cis* chamber unexpectedly shows that this physiological substrate protein is trapped in a unique position and orientation in the chaperonin chamber. The combination of size and shape of the *cis* chamber and gp23 results in the gp23 major domain being wedged into the chamber in a restricted position. The small insertion domain, which makes inter-subunit contacts in the viral capsid<sup>25</sup>, is likely to be mobile and disordered until assembly of the capsid hexamers, which form as soon as the folded gp23 is released from GroEL-gp31 (ref. 22). The encapsulated substrate causes a significant expansion of the folding chamber in the region of the apical domains, and is exerting pressure on the connections between GroEL and gp31 and the inter-subunit contacts between apical domains of the *cis* ring. Thus, even though the folding chamber formed by GroEL-gp31 is larger than that formed by GroEL-GroES, an enclosed gp23 monomer still exerts pressure on it. Previous studies of GroEL-GroES and GroEL-gp31 complexes in various nucleotide states did not reveal any significant changes in the conformation of the folding chamber<sup>20,24</sup> (see also Supplementary Results and Supplementary Discussion). The expansion and compression observed here show that folding substrates directly and indirectly affect the conformation of the chamber. In conclusion, this study reveals a remarkable view of the chaperonin folding chamber strained to encapsulate a physiological substrate protein.



**Figure 4 | Substrate densities isolated from the binary and ternary complexes.** Shown are top and side views of substrate densities isolated from the GroEL-gp23 binary complex (**a**), the GroEL-(gp23)<sub>2</sub> binary complex (**b**), the *trans*-only GroEL-gp23-gp31 ternary complex (**c**) and the *cis* ternary complex (**d**) compared to the low resolution filtered density of the gp24 crystal structure (**e**). The isolated substrate densities were low-pass filtered at 15 Å and their approximate molecular masses were determined at a density threshold of 1σ of the complete complex. A larger observed mass in the class average reflects a more homogeneous class and therefore a more consistent structure for that sub-population. The green line in **e** indicates the major domain of gp24.

## METHODS SUMMARY

The GroEL, gp23 and gp31 were expressed and purified as described<sup>19</sup>. The binary GroEL–gp23 complexes were prepared as described<sup>23</sup>, yielding final concentrations of 1 µM GroEL oligomer, 2.5 µM gp23 monomer and 100 mM urea. GroEL–gp23–gp31 ternary complexes were formed by adding a twofold molar excess of gp31 heptamer over GroEL tetradecamer to preformed binary complexes, along with ADP to give a final concentration of 2.5 mM. After incubation at 24 °C, KF and KAl(SO<sub>4</sub>)<sub>2</sub> were added, giving final concentrations of 20 mM KF and 2 mM KAl(SO<sub>4</sub>)<sub>2</sub>, to form the ATP transition state analogue ADP·AlF<sub>3</sub> to generate folding-active complexes.

Images were recorded on a 200 kV FEG microscope on photographic film and processed at 2.8 Å per pixel, with final data sets of 30,000 and 35,000 side views of the binary and ternary complexes, respectively. A starting model for the binary complex was obtained by angular reconstitution, and our previously determined GroEL–ADP–gp31 structure<sup>20</sup> was used as a starting model for the ternary complexes. The data sets were sorted into classes showing different substrate features by a combination of MSA and competitive projection matching<sup>10</sup>, and the atomic structures of the GroEL subunit domains, gp31 and gp24 subunits were docked into the final, asymmetric maps as separate rigid bodies.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 16 May; accepted 12 September 2008.**

- Houry, W. A. *et al.* Identification of *in vivo* substrates of the chaperonin GroEL. *Nature* **402**, 147–154 (1999).
- Kerner, M. J. *et al.* Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* **122**, 209–220 (2005).
- Rye, H. S. *et al.* GroEL–GroES cycling: ATP and nonnative polypeptide direct alternation of folding-active rings. *Cell* **97**, 325–338 (1999).
- Sigler, P. B. *et al.* Structure and function in GroEL-mediated protein folding. *Annu. Rev. Biochem.* **67**, 581–608 (1998).
- Horwich, A. L., Fenton, W. A., Chapman, E. & Farr, G. W. Two families of chaperonin: Physiology and mechanism. *Annu. Rev. Cell Dev. Biol.* **23**, 115–145 (2007).
- Braig, K. *et al.* The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature* **371**, 578–586 (1994).
- Fenton, W. A., Kashi, Y., Furtak, K. & Horwich, A. L. Residues in chaperonin GroEL required for polypeptide binding and release. *Nature* **371**, 614–619 (1994).
- Lin, Z., Schwartz, F. P. & Eisenstein, E. The hydrophobic nature of GroEL–substrate binding. *J. Biol. Chem.* **270**, 1011–1014 (1995).
- Farr, G. W. *et al.* Multivalent binding of nonnative substrate proteins by the chaperonin GroEL. *Cell* **100**, 561–573 (2000).
- Elad, N. *et al.* Topologies of a substrate protein bound to the chaperonin GroEL. *Mol. Cell* **26**, 415–426 (2007).
- Xu, Z., Horwich, A. L. & Sigler, P. B. The crystal structure of the asymmetric GroEL–GroES–(ADP)<sub>7</sub> chaperonin complex. *Nature* **388**, 741–750 (1997).
- Motojima, F. *et al.* Substrate polypeptide presents a load on the apical domains of the chaperonin GroEL. *Proc. Natl Acad. Sci. USA* **101**, 15005–15012 (2004).
- Lin, Z. & Rye, H. S. Expansion and compression of a protein folding intermediate by GroEL. *Mol. Cell* **16**, 23–34 (2004).
- Lin, Z., Madan, D. & Rye, H. S. GroEL stimulates protein folding through forced unfolding. *Nature Struct. Mol. Biol.* **15**, 303–311 (2008).
- Mayhew, M. *et al.* Protein folding in the central cavity of the GroEL–GroES chaperonin complex. *Nature* **379**, 420–426 (1996).
- Laemmli, U. K., Beguin, F. & Gujer-Kellenberger, G. A factor preventing the major head protein of bacteriophage T4 from random aggregation. *J. Mol. Biol.* **47**, 69–85 (1970).
- van der Vies, S. M., Gatenby, A. A. & Georgopoulos, C. Bacteriophage T4 encodes a co-chaperonin that can substitute for *Escherichia coli* GroES in protein folding. *Nature* **368**, 654–656 (1994).
- Hunt, J. F., van der Vies, S. M., Henry, L. & Deisenhofer, J. Structural adaptations in the specialized bacteriophage T4 co-chaperonin Gp31 expand the size of the Anfinsen cage. *Cell* **90**, 361–371 (1997).
- Bakkes, P. J., Faber, B. W., van Heerikhuizen, H. & van der Vies, S. M. The T4-encoded cochaperonin, gp31, has unique properties that explain its requirement for the folding of the T4 major capsid protein. *Proc. Natl Acad. Sci. USA* **102**, 8144–8149 (2005).
- Clare, D. K. *et al.* An expanded protein folding cage in the GroEL–gp31 complex. *J. Mol. Biol.* **358**, 905–911 (2006).
- Elad, N., Clare, D. K., Saibil, H. R. & Orlova, E. V. Detection and separation of heterogeneity in molecular complexes by statistical analysis of their two-dimensional projections. *J. Struct. Biol.* **162**, 108–120 (2008).
- van Duijn, E. *et al.* Monitoring macromolecular complexes involved in the chaperonin-assisted protein folding cycle by mass spectrometry. *Nature Methods* **2**, 371–376 (2005).
- van Duijn, E., Heck, A. J. & van der Vies, S. M. Inter-ring communication allows the GroEL chaperonin complex to distinguish between different substrates. *Protein Sci.* **16**, 956–965 (2007).
- Ranson, N. A. *et al.* Allosteric signaling of ATP hydrolysis in GroEL–GroES complexes. *Nature Struct. Mol. Biol.* **13**, 147–152 (2006).
- Fokine, A. *et al.* Structural and functional similarities between the capsid proteins of bacteriophages T4 and HK97 point to a common ancestry. *Proc. Natl Acad. Sci. USA* **102**, 7163–7168 (2005).
- van Duijn, E. *et al.* Tandem mass spectrometry of intact GroEL–substrate complexes reveals substrate-specific conformational changes in the trans ring. *J. Am. Chem. Soc.* **128**, 4694–4702 (2006).
- Snyder, L. & Tarkowski, H. J. The N terminus of the head protein of T4 bacteriophage directs proteins to the GroEL chaperonin. *J. Mol. Biol.* **345**, 375–386 (2005).
- Falke, S. *et al.* The 13 angstroms structure of a chaperonin GroEL–protein substrate complex by cryo-electron microscopy. *J. Mol. Biol.* **348**, 219–230 (2005).
- Tang, Y. C. *et al.* Structural features of the GroEL–GroES nano-cage required for rapid folding of encapsulated protein. *Cell* **125**, 903–914 (2006).
- Machida, K. *et al.* Hydrophilic residues 526KNDAD531 in the flexible C-terminal region of the chaperonin GroEL are critical for substrate protein folding within the central cavity. *J. Biol. Chem.* **283**, 6886–6896 (2008).
- Chaudhry, C. *et al.* Role of the gamma-phosphate of ATP in triggering protein folding by GroEL–GroES: Function, structure and energetics. *EMBO J.* **22**, 4877–4887 (2003).
- Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
- Crowther, R. A., Henderson, R. & Smith, J. M. MRC image processing programs. *J. Struct. Biol.* **116**, 9–16 (1996).
- Frank, J. *et al.* SPIDER and WEB: Processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
- van Heel, M. *et al.* A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116**, 17–24 (1996).
- Navaza, J. *et al.* On the fitting of model electron densities into EM reconstructions: A reciprocal-space formulation. *Acta Crystallogr. D* **58**, 1820–1825 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank R. Westlake, D. Houldershaw and L. Wang for computing and EM support. This work was carried out at the School of Crystallography, Birkbeck College, and was supported by a Wellcome Trust programme grant, EU 3D EM Network of Excellence and 3D Repertoire grants.

**Author Information** The three-dimensional reconstructions for the two binary complexes and the three ternary complexes are deposited in the EBI-MSD EMD database with accession codes EMD-1544 through to EMD-1548. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to H.R.S. (h.saibil@mail.cryst.bbk.ac.uk) and S.M.v.d.V. (vdvies@vumc.nl).



## METHODS

**Sample preparation. Binary complexes.** GroEL was diluted to a concentration of 1.2  $\mu\text{M}$  (oligomer) in 40 mM Tris-HCl pH 7.5, 10 mM KCl, 10 mM  $\text{MgCl}_2$ . 25  $\mu\text{M}$  gp23 monomer was denatured in 6 M urea for 2 h at 24 °C and 300 r.p.m. on a bench top shaker. The denatured gp23 was then added to 90  $\mu\text{l}$  of GroEL solution in ten 1- $\mu\text{l}$  additions, with vortexing and centrifugation after each addition, giving a final concentration of 1  $\mu\text{M}$  GroEL, 2.5  $\mu\text{M}$  gp23 and 600 mM urea. To maximize the occupancy of the gp23–GroEL bound complex, the denatured gp23 was added in a 2.5 molar excess over GroEL oligomer. The urea was diluted to 100 mM by addition of 500  $\mu\text{l}$  of buffer to the GroEL–gp23 complex. This mixture was then concentrated in a Vivaspin with a 5 kDa cut-off, to a final concentration of 1  $\mu\text{M}$  GroEL oligomer, 2.5  $\mu\text{M}$  gp23 monomer and 100 mM urea.

**Ternary complexes.** The non-hydrolysable ATP analogue ADP· $\text{AlF}_3$  has been shown to produce folding active ternary complexes for several GroEL substrates<sup>31</sup>. We therefore used it to produce folding-active GroEL–gp23–gp31:ADP· $\text{AlF}_3$  complexes. 0.75  $\mu\text{l}$  of 140  $\mu\text{M}$  (heptamer) gp31 was added to 50  $\mu\text{l}$  of binary complex prepared as above, followed by 0.3  $\mu\text{l}$  of 400 mM ADP. The mixture was incubated for 10 min at 24 °C. Then 1  $\mu\text{l}$  of 1 M KF was added and mixed, followed by 1  $\mu\text{l}$  of 100 mM  $\text{KAl}(\text{SO}_4)_2$ , mixing and 5 min incubation. To maximize the number of ternary complexes, a twofold molar excess of gp31 over GroEL was added to the preformed binary complexes.

**Data collection.** Samples (3.5  $\mu\text{l}$ ) were loaded on to freshly glow discharged (45 s) c2/2 holey carbon C-flat grids (Protochips Inc.), blotted and plunged into liquid nitrogen cooled liquid ethane. Low dose images were recorded on an F20 microscope (FEI) at 50,000 $\times$  on Kodak SO-163 film with a defocus range of 1–3  $\mu\text{m}$ .

The images showed an even distribution of side and end views very similar to that seen for GroEL alone. The number of side views is crucial to the study of GroEL complexes as they contain all the information required to reconstruct the molecule in three dimensions. In contrast, for the GroEL–MDH binary complex, a modified GroEL had to be used to increase the number of side views<sup>10</sup>.

**Image processing.** The images were digitized at a step size of 7  $\mu\text{m}$  on a Zeiss photodensitometer (Intergraph), giving a sampling of 1.4 Å per pixel. The defocus was determined with CTFFIND3<sup>32</sup>. Particles were manually picked with XIMDISP and extracted into 512  $\times$  512 pixel boxes with LABEL<sup>33</sup>. The cut-out particles were corrected for the phase reversals of the contrast transfer function (CTF) in SPIDER<sup>34</sup>, cropped to 320  $\times$  320 pixels and 2  $\times$  2 averaged to give a final sampling of 2.8 Å per pixel. The boxed particles were filtered between 200 and 6 Å for the binary complexes and between 285 and 6 Å for the ternary complexes and then normalized. The images were centred in SPIDER using side and end view projections of the crystal structures of GroEL and GroEL–GroES that were low pass filtered to 40 Å. The aligned images were then classified using MSA in IMAGIC into classes each containing on average five images<sup>35</sup>. Using these classes, end views, bad images and GroEL lacking gp31 in the ternary complex sample were removed to produce two data sets containing 30,000 and 35,000 side view images of the binary and ternary complexes, respectively. The binary data set, including an end view class, was aligned and initially reconstructed by angular reconstitution in IMAGIC with imposed seven-fold symmetry, treating it as a single structure.

**Binary complexes.** The initial map was used as a starting model for projection matching in SPIDER, with the number of references increasing to 155 (90–105° in  $\beta$  and 0–51.4° in  $\gamma$ ) which gave a  $\gamma$  spacing of 2.1°. The resulting reconstruction was masked to generate two reference structures, one without substrate density and the other containing one-third of the sevenfold averaged substrate density. At this stage the symmetry was relaxed from *C*<sub>7</sub> to *C*<sub>1</sub> and the reconstructions were used for competitive projection matching, in which the images are sorted by cross-correlation value to projections of both reference structures. The images

that aligned to the substrate-bound reference were then subjected to MSA, and two eigenimages that represented substrate variance were used to separate the images into two populations of substrate-bound structures, GroEL–gp23 and GroEL–(gp23)<sub>2</sub>, which, along with the empty structure, were used for further competitive projection matching.

For the asymmetric refinement, the number of reference images was increased from 155 to 528 per reconstruction for further competitive alignment with an angular spacing of 3° from 78–90° in  $\beta$  and 3.1° from 0–360° in  $\gamma$  and with the mirror option turned on in the SPIDER operation AP SH. After another three rounds of alignment, when the image alignments were stable for each of the three species, they were subjected to MSA. This led to a further separation to give one empty, two GroEL–gp23 and two GroEL–(gp23)<sub>2</sub> species. The separation was refined again by projection matching. When the alignment had stabilized, 95% of the images aligned to the same references in consecutive alignments. MSA was carried out on each of the subsets to check that the substrate variations had been minimized (Supplementary Fig. 3a–c). The total number of images in the data set limited the number of different complexes that could be discriminated and therefore no more separations were carried out. The resolutions of the five final reconstructions, determined by the 0.5 criterion of the Fourier shell correlation, were 11–12 Å.

**Ternary complexes.** The initial model used to align the ternary complexes was the previously published GroEL–gp31–ADP structure<sup>20</sup>, re-projected at an angular spacing of 2.1° (90–110° in  $\beta$  and 0–51.4° in  $\gamma$ ). The resulting seven-fold symmetric reconstruction contained density for both *cis* and *trans* bound substrate and was used for further alignment and MSA, eventually yielding five classes. Classes 1–3 together represented 70% of the data and contained well aligned images, whereas classes 4 and 5 contained poorly aligned images. After this separation we continued the alignment procedure with classes 1, 2 and 3 as independent data sets. The alignment of these three classes stabilized quickly with 95% of the images aligning to the same reference in consecutive rounds, and the calculated eigenimages showing that the variation due to substrate had been removed (Supplementary Fig. 3d–f).

Once the alignment had stabilized for each of the complexes the symmetry was relaxed from *C*<sub>7</sub> to *C*<sub>1</sub> and the number of references increased to give a final angular spacing of 2° (90–110° in  $\beta$  and 0–360° in  $\gamma$  and with the mirror option turned on in the SPIDER operation AP SH). The refinement was stopped when more than 95% of the images aligned to same reference in consecutive rounds of alignment. The images of the *cis/trans* bound complex were also aligned to the *trans* only bound references and the resulting reconstruction had almost identical *cis* substrate density, confirming the reliability of the *cis* substrate density. The resolutions of the three maps were all around 10 Å by the 0.5 Fourier shell correlation criterion.

**Docking of atomic coordinates.** As there was no symmetry applied to the reconstructions, all 14 subunits of GroEL (1GRL and 1AON) and 7 subunits of gp31 (1G31) were fitted individually. Each of the GroEL subunits was further split into its three separate domains, apical 192–373, intermediate 137–191 and 373–409, and equatorial 2–136 and 410–525, which were docked as rigid bodies. The mobile loops of gp31 (residues 15–36) were omitted since they were not in the GroEL-bound conformation in the crystal structure. The initial docking was done manually in Pymol (www.pymol.org) and then refined by the automatic docking program URO<sup>36</sup>. However the fitting of the intermediate domains was not accurate since the density for these domains in most of the maps was not as well defined as that of the other domains. Therefore, the intermediate domains had to be checked and manually corrected to maintain the connectivity between the equatorial and intermediate domains. After adjustment of the intermediate domains all the connections between it and the equatorial and apical domains were within 7 Å ( $\alpha$ -carbon distance), with the majority less than that.

# X-ray structure of a pentameric ligand-gated ion channel in an apparently open conformation

Nicolas Bocquet<sup>1\*</sup>, Hugues Nury<sup>1,2\*</sup>, Marc Baaden<sup>4</sup>, Chantal Le Poupon<sup>1</sup>, Jean-Pierre Changeux<sup>3</sup>, Marc Delarue<sup>2</sup> & Pierre-Jean Corringer<sup>1</sup>

Pentameric ligand-gated ion channels from the Cys-loop family mediate fast chemo-electrical transduction<sup>1–3</sup>, but the mechanisms of ion permeation and gating of these membrane proteins remain elusive. Here we present the X-ray structure at 2.9 Å resolution of the bacterial *Gloeobacter violaceus* pentameric ligand-gated ion channel homologue<sup>4</sup> (GLIC) at pH 4.6 in an apparently open conformation. This cationic channel is known to be permanently activated by protons<sup>5</sup>. The structure is arranged as a funnel-shaped transmembrane pore widely open on the outer side and lined by hydrophobic residues. On the inner side, a 5 Å constriction matches with rings of hydrophilic residues that are likely to contribute to the ionic selectivity<sup>6–9</sup>. Structural comparison with ELIC, a bacterial homologue from *Erwinia chrysanthemi* solved in a presumed closed conformation<sup>10</sup>, shows a wider pore where the narrow hydrophobic constriction found in ELIC is removed. Comparative analysis of GLIC and ELIC reveals, in concert, a rotation of each extracellular  $\beta$ -sandwich domain as a rigid body, interface rearrangements, and a reorganization of the transmembrane domain, involving a tilt of the M2 and M3  $\alpha$ -helices away from the pore axis. These data are consistent with a model of pore opening based on both quaternary twist and tertiary deformation.

Pentameric ligand-gated ion channels (pLGICs) are allosteric proteins regulating cellular excitability through the opening of an intrinsic transmembrane ion channel. Agonist binding to extracellular sites shifts a closed conformation into an open one, allowing ions to diffuse down their electrochemical gradient. Two agonist-free structures have been reported: the *Torpedo marmorata* nicotinic acetylcholine receptor (nAChR-*Tm*), solved by electron microscopy at 4 Å resolution<sup>11,12</sup>, and the 3.3 Å ELIC X-ray structure<sup>10</sup>. X-ray structures of homologues of the extracellular domain (ECD) of nAChRs have also been described: the acetylcholine-binding proteins (AChBPs) co-crystallized with agonists and antagonists<sup>13–15</sup>, and the ECD of  $\alpha 1$ -nAChR<sup>16</sup>. A key issue is now to understand how pLGICs open, select and translocate ions through the membrane. Most pLGICs undergo desensitization on prolonged exposure to agonist, complicating structural investigations of the transient open conformation. In contrast, GLIC is activated by protons but does not desensitize, even at proton concentrations eliciting the maximal electrophysiological response (pH 4.5)<sup>5</sup>. Here we present the first apparently open structure of this family, GLIC crystallized at pH 4.6.

The overall architecture of GLIC is similar to those of ELIC, the AChBPs, and nAChR-*Tm* (Fig. 1a). The five subunits are arranged in a barrel-like manner around a central symmetry axis that coincides with the ion permeation pathway (Fig. 1b). Subunits interact tightly through a 2,200 Å<sup>2</sup> interface containing charged residues and water molecules (Supplementary Fig. 8). The ECD of each subunit consists

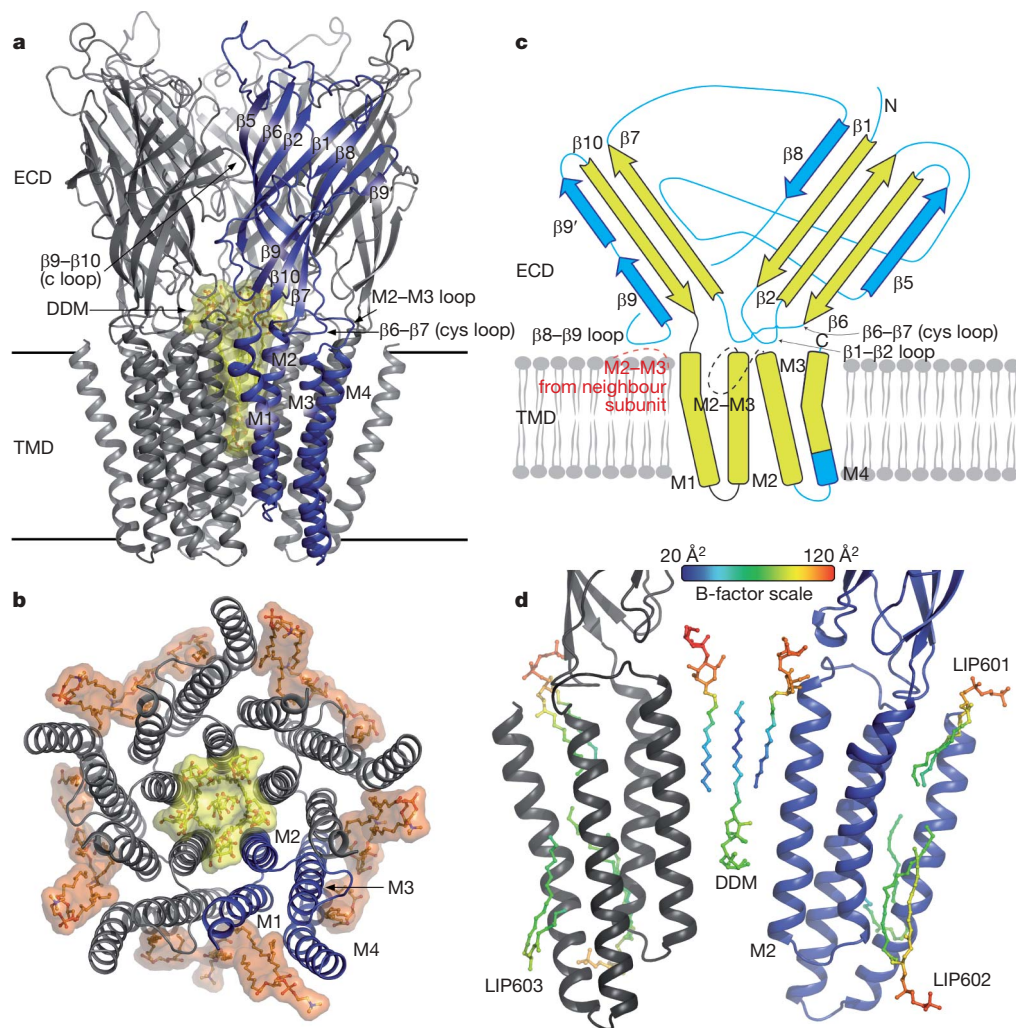
of a  $\beta$ -sandwich composed of five inner and three outer strands connected by loops (Fig. 1c). In eukaryotic pLGICs, the interface between ECDs holds the neurotransmitter-binding pocket<sup>2,14</sup>. The equivalent region in GLIC presents structural similarity despite a low sequence identity, and is well defined, notably the capping  $\beta 9$ – $\beta 10$  loop (loop C)<sup>15</sup>. The transmembrane domain (TMD) of each subunit consists of four helices (M1 to M4). M2 helices form the wall of the pore (Fig. 2a), bordered by rings of homologous residues, as previously established in nAChRs<sup>6,17</sup>: a charged E–2', two polar T2', S6', and three hydrophobic I9', A13' and I16'/L17' rings, that are close/homologous to  $\alpha 1$ -nAChR-*Tm* E–1', T2', S6', L9', V13' and L16'/V17' (the prime numbering starts at approximately the beginning of M2, at positions homologous to  $\alpha 1$ K242 of nAChR-*Tm*, Fig. 2c). M1 helices are kinked at P205 and form with M3 a second circle of helices interacting with M2. M4 helices are peripheral. Well-defined electron densities are observed in the grooves between M4 and both M1 and M3, close to residues labelled by hydrophobic probes on nAChR-*Tm*<sup>18</sup>. They were attributed to lipids that possibly contribute to holding M4 in its position (Fig. 1b, d, and Supplementary Fig. 3). The central ion permeation pathway consists of an extracellular hydrophilic vestibule more than 12 Å wide, followed by a funnel-shaped transmembrane pore (Fig. 2d). The M2 axes are tilted with respect to the pore axis, with outer hydrophobic side chains oriented towards the helix interfaces, and inner polar side chains oriented towards the pore.

A bundle of six detergent molecules (dodecyl- $\beta$ -D-maltoside, DDM) obstructs the pore, with one detergent per monomer and one sitting on the five-fold axis, the sugar moiety being much less ordered than the aliphatic chains (Fig. 1b, d, Supplementary Fig. 3). The former interact extensively with the hydrophobic rings, shielding their side chains from the solvent with the polar heads pointing up, towards the vestibule, while the latter points its polar head down. To probe the influence of the detergent on the protein conformation, diffraction data have been collected with crystals grown in the presence of two DDM analogues (Supplementary Fig. 4). These detergents contain two bulky bromine atoms in their alkyl chain and could not bind all in the same position as described here (Fig. 1) without seriously disturbing the pore structure, for steric reasons. In both cases we observe that the protein conformation is unchanged (r.m.s. deviation <0.3 Å), whereas either the central detergent is not seen in the electron density (10,11-dibromoundecanoyl- $\beta$ -maltoside) or the five central detergent-binding sites are only partly occupied (7,8-dibromododecyl- $\beta$ -maltoside). Furthermore, we performed molecular dynamics simulations (20 ns) of the GLIC molecule in a lipid bilayer without DDM in the pore. We observe that its conformation is stable throughout the simulation (Supplementary Fig. 6).

<sup>1</sup>Pasteur Institute, G5 Group of Channel-Receptor, CNRS URA 2182. <sup>2</sup>Pasteur Institute, Unit of Structural Dynamics of Macromolecules, CNRS URA 2185. <sup>3</sup>Pasteur Institute, CNRS URA 2182, F75015, Paris, France. <sup>4</sup>Institut de Biologie Physico-Chimique, CNRS UPR 9080, 75005 Paris, France.

\*These authors contributed equally to this work.





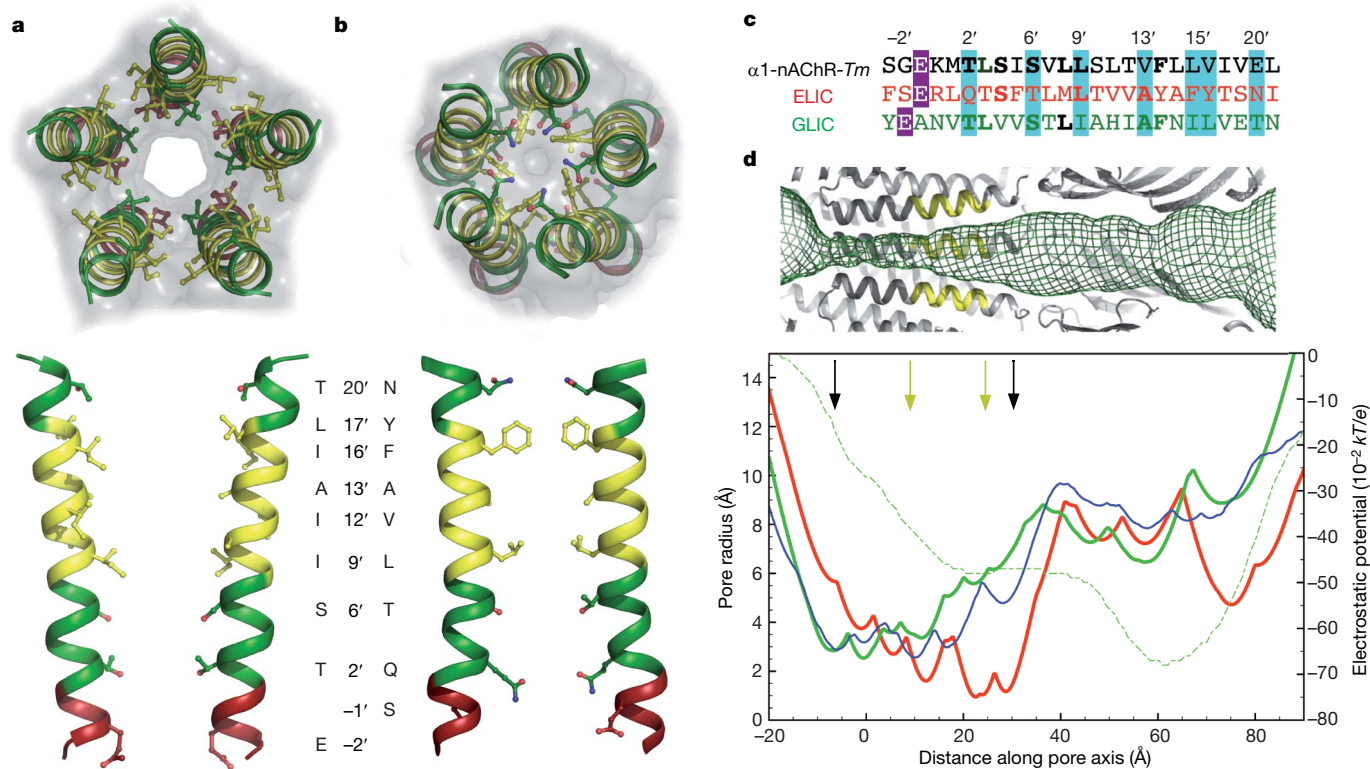
**Figure 1 | GLIC structure.** **a**, Ribbon representation of GLIC viewed from the plane of the membrane. DDM molecules bound in the channel are depicted as yellow sticks plus van der Waals surface. Horizontal lines represent the membrane limits. **b**, Transmembrane part of GLIC viewed from the extracellular side. The ECD is removed for clarity. Lipids are also

depicted in orange. **c**, Topology of a GLIC subunit. The conserved core elements common to GLIC and ELIC are coloured in yellow. **d**, Close-up view of the TMD. Only two subunits are represented. The DDM molecules and the lipids (named LIP601/2/3) close to these subunits are coloured according to their atomic B-factor (colour scale at top).

The pore of GLIC resembles a cone with 12 Å outside and 5 Å inside diameters, sharply contrasting with that of ELIC, where outer residues L9', A13' and F16' point their side chains towards the central axis, creating a 2 Å diameter hydrophobic barrier expected to prevent ion permeation (Fig. 2b). This wide opening of the pore is suggestive of an open structure. Three lines of evidence support this idea. First, GLIC was solved in the presence of a non-desensitizing agonist. Second, the GLIC pore structure is consistent with a wealth of biochemical data obtained on the open conformation of nAChRs that located the channel constriction between residues -2' and 2' (refs 19, 20), from measurement of cysteine accessibility<sup>19</sup> and channel conductance decreases following positively charged residue substitution<sup>20</sup>. Yet, permeability studies with organic cations suggest that this constriction is 7.5 Å wide in nAChRs<sup>8</sup>, while it is only 5 Å wide in GLIC. Intrinsic flexibility of the side chains composing the constriction could compensate for this difference and transiently allow for accommodation of large cations. Third, molecular dynamics simulations based on the nAChR-*Tm* confirm that it corresponds to an impermeant structure, with rings 9' and 13' creating a hydrophobic barrier, but show that small increases (less than 2 Å) in diameter would produce a permeant channel<sup>21,22</sup>. The GLIC diameter is significantly wider than that of nAChR-*Tm* at these positions, further highlighting its compatibility with cation permeation.

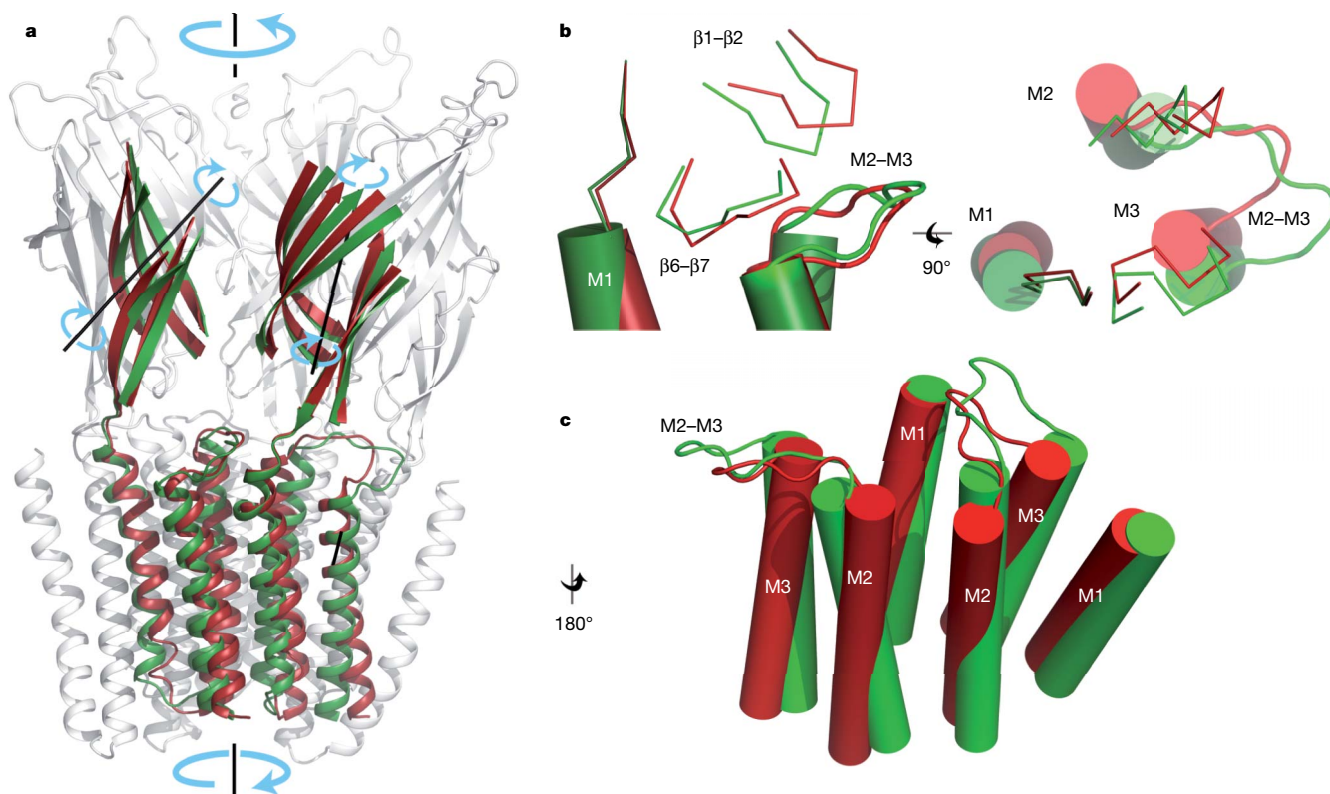
Interestingly, mutagenesis has pointed to a key role of the -2'/2' region in determining charge selectivity<sup>6-9</sup>; a ring of negatively charged residues at -1' and an additional ring of prolines at -2' respectively favour cationic and anionic selectivity of eukaryotic pLGICs. Thus, the -2'/2' region was proposed to constitute the selectivity filter of the pore<sup>7-9</sup>, an idea consistent with our structure, where the constriction is formed by the T2' hydroxyl moieties, flanked on both sides by the S6' hydroxyl and E-2' carboxyl moieties. The 5 Å diameter at T2' is too narrow to fit a fully hydrated ion, as indicated by the average distance between sodium or potassium and protein side-chain oxygen atoms (respectively 2.4 and 2.85 Å between atom centres<sup>23</sup>). Cation permeation would require the loss of equatorial water molecules; yet rings of polar side chains are located in the right position to transiently coordinate cations and could compensate for this loss. Regarding the pore conformation, data are thus consistent with the notion that GLIC and ELIC structures are in presumed open and closed conformations, respectively, while that of nAChR-*Tm* might be in a functionally closed but structurally nearly open conformation.

Assuming that GLIC and ELIC structures faithfully represent the open and closed forms of pLGICs embedded in a phospholipid bilayer, insight into the opening mechanism can be provided by analysing their rearrangements, even though they share only 18%



**Figure 2 | GLIC and ELIC pores.** **a**, Top and side view of GLIC M2 helices (top and bottom panels, respectively); helix backbones and side chains facing the pore are depicted. Hydrophobic, polar and negative residues are coloured yellow, green and red respectively. In the side view, only two subunits are shown. DDM molecules are removed for clarity. **b**, Top and side views of ELIC

M2 helices. **c**, M2 sequence alignment. **d**, Pore radius for  $\alpha 1$ -nAChR-Tm (blue), ELIC (red) and GLIC (green) along its axis. The upper part is a mesh representation at the same scale of the GLIC channel. The electrostatic potential is represented as a dotted green line. Black and yellow arrows indicate the limits of the membrane and the hydrophobic rings, respectively.



**Figure 3 | Open GLIC and closed ELIC structure comparison.** **a**, Side view of the structural superposition. For the two subunits in the foreground, only the common core is depicted, in green for GLIC, in red for ELIC. Other subunits are in grey. The ECD rotation axes and the twist axis are depicted.

The M4 helix is omitted for clarity. **b**, Close-up view of the interface between the ECD and the TMD (side view in left panel and upper view in right panel). **c**, Close-up of transmembrane helices M1–M3 viewed from the channel.



sequence identity. Several elements can be unambiguously aligned with no gap or insertion, with identical secondary structure (Supplementary Fig. 2): helices M1, M2 and M3, and a large portion of the  $\beta$ -sandwich consisting of strands  $\beta$ 1,  $\beta$ 2,  $\beta$ 6,  $\beta$ 7 and  $\beta$ 10, also well conserved in the AChBPs. These elements constitute the subunit 'common core'. Common core superimposition shows that the GLIC subunits display a quaternary twist compared to ELIC, with anti-clockwise (versus clockwise) rotation in the upper (versus lower) part of the pentamer, when viewed from the extracellular compartment (Fig. 3a). This is confirmed by normal mode analysis: the lowest frequency mode is precisely a twist mode and has by far the highest contribution (29%) to the transition (Supplementary Fig. 7a). However, we note that the first 100 lowest-frequency modes (usually the most collective ones) only explain about 50% of the transition. Other and more local movements occur: in the TMD, the outer ends of M2 and M3 of GLIC are tilted away radially from the channel axis, while the outer end of M1 is fixed. The inner ends of M1, M2 and M3 move tangentially towards the left, when viewed from the membrane (Fig. 3b). In the ECD, the core of the  $\beta$ -sandwich undergoes little deformation (Supplementary Table 2), but is rotated by  $8^\circ$  around an axis roughly perpendicular to the inner sheet of the  $\beta$ -sandwich (Fig. 3a), concomitant with a rearrangement of both the subunit-subunit and the ECD/TMD interfaces, regions known to contribute to neurotransmitter gating<sup>24–27</sup>. The latter contains the well-conserved  $\beta$ 6– $\beta$ 7 and M2–M3 loops and the  $\beta$ 1– $\beta$ 2 loop whose length is conserved in the pLGIC family. We observe a downward motion of the  $\beta$ 1– $\beta$ 2 loop, concomitant with a displacement of the M2–M3 loop, M2 and M3 helices and  $\beta$ 6– $\beta$ 7 loop towards the periphery of the molecule (Fig. 3c), thereby opening the pore.

This mechanism is different from a more local gating pathway suggested by the comparison of the conformations of the  $\alpha$  and non- $\alpha$  subunits in nAChR-*Tm*<sup>12</sup>. It implies both a quaternary twist and tertiary deformations. Such twist to open motions, initially proposed from *ab initio* normal mode analysis of nAChRs<sup>28</sup>, and observed for KcsA<sup>29</sup>, may plausibly be extended to eukaryotic pLGICs. The structural transition described here couples in an allosteric manner the opening-closing motion of the pore with distant binding sites—located at the ECD subunit interface for neurotransmitters, or within the TMD for allosteric effectors<sup>30</sup>—and may possibly serve as a general mechanism of signal transduction in pLGICs.

## METHODS SUMMARY

GLIC was expressed as described previously<sup>5</sup>, and solubilized/purified in 2%/0.02% DDM including an amylose affinity chromatography, two size exclusion chromatographies and thrombin cleavage of fused MBP. Crystals were grown at pH 4.6, with 15% PEG 4000, 100 mM NaOAc and 400 mM (NH<sub>4</sub>)SCN in hanging drops. Data were collected at MX beamlines of the European Synchrotron Radiation Facility, and processed with XDS and CCP4 programs (Supplementary Table 1). The spacegroup is C2 with one pentamer in the asymmetric unit (Supplementary Fig. 1). Molecular replacement with Phaser using the ELIC structure gave the first density maps. Refinement was conducted with Coot and Refmac, using tight NCS restraints. The final model presents a good geometry and consists of residues 6–315 for the 5 subunits plus 6 DDM and 15 partial lipid molecules, and 115 water molecules (Supplementary Fig. 3).

Received 2 June 2008; accepted 29 September 2008.

Published online 5 November 2008.

- Corringer, P. J. & Changeux, J. P. Nicotinic acetylcholine receptors. *Scholarpedia* **3**, 3468 (2008).
- Lester, H. A., Dibas, M. I., Dahan, D. S., Leite, J. F. & Dougherty, D. A. Cys-loop receptors: New twists and turns. *Trends Neurosci.* **6**, 329–336 (2004).
- Sine, S. M. & Engel, A. G. Recent advances in Cys-loop receptor structure and function. *Nature* **440**, 448–455 (2006).
- Tasneem, A., Iyer, L. M., Jakobsson, E. & Aravind, L. Identification of the prokaryotic ligand-gated ion channels and their implications for the mechanisms and origins of animal Cys-loop ion channels. *Genome Biol.* **6**, R4 (2005).
- Bocquet, N. *et al.* A prokaryotic proton-gated ion channel from the nicotinic acetylcholine receptor family. *Nature* **445**, 116–119 (2007).
- Imoto, K. *et al.* Rings of negatively charged amino acids determine the acetylcholine receptor channel conductance. *Nature* **335**, 645–648 (1988).

- Imoto, K. *et al.* A ring of uncharged polar amino acids as a component of channel constriction in the nicotinic acetylcholine receptor. *FEBS Lett.* **289**, 193–200 (1991).
- Keramidas, A., Moorhouse, A. J., Schofield, P. R. & Barry, P. H. Ligand-gated ion channels: Mechanisms underlying ion selectivity. *Prog. Biophys. Mol. Biol.* **86**, 161–204 (2004).
- Corringer, P. J. *et al.* Mutational analysis of the charge selectivity filter of the  $\alpha$ 7 nicotinic acetylcholine receptor. *Neuron* **22**, 831–843 (1999).
- Hilf, R. J. & Dutzler, R. X-ray structure of a prokaryotic pentameric ligand-gated ion channel. *Nature* **452**, 375–379 (2008).
- Miyazawa, A., Fujiyoshi, Y. & Unwin, N. Structure and gating mechanism of the acetylcholine receptor pore. *Nature* **423**, 949–955 (2003).
- Unwin, N. Refined structure of the nicotinic acetylcholine receptor at 4 Å resolution. *J. Mol. Biol.* **346**, 967–989 (2005).
- Brej, K. *et al.* Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature* **411**, 269–276 (2001).
- Celie, P. H. *et al.* Nicotine and carbamylcholine binding to nicotinic acetylcholine receptors as studied in AChBP crystal structures. *Neuron* **41**, 907–914 (2004).
- Hansen, S. B. *et al.* Structures of *Aplysia* AChBP complexes with nicotinic agonists and antagonists reveal distinctive binding interfaces and conformations. *EMBO J.* **24**, 3635–3646 (2005).
- Dellisanti, C. D., Yao, Y., Stroud, J. C., Wang, Z. Z. & Chen, L. Crystal structure of the extracellular domain of nAChR  $\alpha$ 1 bound to  $\alpha$ -bungarotoxin at 1.94 Å resolution. *Nature Neurosci.* **10**, 953–962 (2007).
- Giraudat, J., Dennis, M., Heidmann, T., Chang, J. Y. & Changeux, J. P. Structure of the high-affinity binding site for noncompetitive blockers of the acetylcholine receptor: Serine-262 of the delta subunit is labeled by [<sup>3</sup>H]chlorpromazine. *Proc. Natl Acad. Sci. USA* **83**, 2719–2723 (1986).
- Blanton, M. P. & Cohen, J. B. Identifying the lipid-protein interface of the *Torpedo* nicotinic acetylcholine receptor: Secondary structure implications. *Biochemistry* **33**, 2859–2872 (1994).
- Wilson, G. G. & Karlin, A. The location of the gate in the acetylcholine receptor channel. *Neuron* **20**, 1269–1281 (1998).
- Cymes, G. D., Ni, Y. & Grosman, C. Probing ion-channel pores one proton at a time. *Nature* **438**, 975–980 (2005).
- Wang, H. L., Cheng, X., Taylor, P., McCammon, J. A. & Sine, S. M. Control of cation permeation through the nicotinic receptor channel. *PLOS Comput. Biol.* **4**, e41 (2008).
- Beckstein, O. & Sansom, M. S. The influence of geometry, surface character, and flexibility on the permeation of ions and water through biological pores. *Phys. Biol.* **1**, 42–52 (2004).
- Harding, M. M. Metal-ligand geometry relevant to proteins and in proteins: Sodium and potassium. *Acta Crystallogr. D* **58**, 872–874 (2002).
- Jha, A., Cadugan, D. J., Purohit, P. & Auerbach, A. Acetylcholine receptor gating at extracellular transmembrane domain interface: The cys-loop and M2–M3 linker. *J. Gen. Physiol.* **130**, 547–558 (2007).
- Grutter, T. *et al.* Molecular tuning of fast gating in pentameric ligand-gated ion channels. *Proc. Natl Acad. Sci. USA* **102**, 18207–18212 (2005).
- Lee, W. Y. & Sine, S. M. Principal pathway coupling agonist binding to channel gating in nicotinic receptors. *Nature* **438**, 243–247 (2005).
- Lummis, S. C. *et al.* Cis-trans isomerization at a proline opens the pore of a neurotransmitter-gated ion channel. *Nature* **438**, 248–252 (2005).
- Taly, A. *et al.* Normal mode analysis suggests a quaternary twist model for the nicotinic receptor gating mechanism. *Biophys. J.* **88**, 3954–3965 (2005).
- Shimizu, H. *et al.* Global twisting motion of single molecular KcsA potassium channel upon gating. *Cell* **132**, 67–78 (2008).
- Li, G. D. *et al.* Identification of a GABAA receptor anesthetic binding site at subunit interfaces by photolabeling with an etomidate analog. *J. Neurosci.* **26**, 11599–11605 (2006).

Supplementary Information is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank P. Koehl for his program Aquasol and help with electrostatic calculations; P. Delepelaire and S. Edelstein for discussions; the staff of ESRF (Grenoble) ID14 and ID23 beamlines for data collection; facilities of the Pasteur Institute (A. Haouz for crystallography, P. England for ultracentrifugation experiments, J. d'Alayer for mass spectroscopy controls and J. Bellalou for help in protein expression); and B. De Foresta (CEA, Orsay) for a gift of the two brominated DDM analogues. The latter diffraction data sets were collected at SLS and PSI (Villingen, Switzerland). We thank M. Fuchs for assistance during data collection; and the IDRIS supercomputer centre and its support staff for allocating CPU time at very short notice (project 082292). This work was supported by the Région Ile-de-France (N.B.), the Association Française contre les Myopathies, the Collège de France (C.L.P.), the Commission of the European Communities (Neurocyprines project; H.N.) and the Network of European Neuroscience Institutes (ENI-NET).

**Author Information** Coordinates of GLIC have been deposited in the Protein Data Bank under accession number 3eam. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to M.D. ([marc.delarue@pasteur.fr](mailto:marc.delarue@pasteur.fr)) or P.-J.C. ([pjcorrin@pasteur.fr](mailto:pjcorrin@pasteur.fr)).

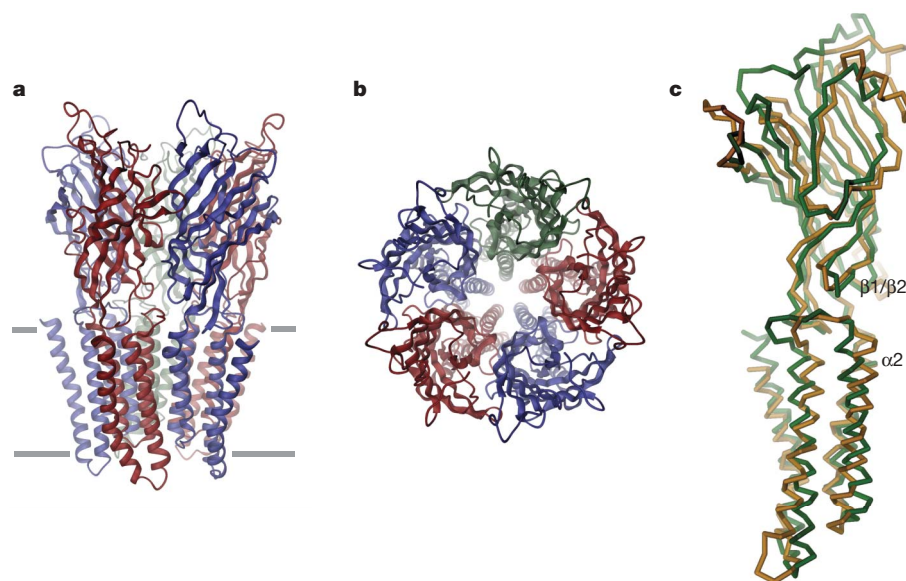
# Structure of a potentially open state of a proton-activated pentameric ligand-gated ion channel

Ricarda J. C. Hilf<sup>1</sup> & Raimund Dutzler<sup>1</sup>

The X-ray structure of a pentameric ligand-gated ion channel from *Erwinia chrysanthemi* (ELIC) has recently provided structural insight into this family of ion channels at high resolution<sup>1</sup>. The structure shows a homo-pentameric protein with a barrel-stave architecture that defines an ion-conduction pore located on the fivefold axis of symmetry. In this structure, the wide aqueous vestibule that is encircled by the extracellular ligand-binding domains of the five subunits narrows to a discontinuous pore that spans the lipid bilayer. The pore is constricted by bulky hydrophobic residues towards the extracellular side, which probably serve as barriers that prevent the diffusion of ions. This interrupted pore architecture in ELIC thus depicts a non-conducting conformation of a pentameric ligand-gated ion channel, the thermodynamically stable state in the absence of bound ligand. As ligand binding promotes pore opening in these ion channels and the specific ligand for ELIC has not yet been identified, we have turned our attention towards a homologous protein from the cyanobacterium *Gloeobacter violaceus* (GLIC). GLIC was shown to form proton-gated channels that are activated by a pH decrease on the extracellular side and that do not desensitize after activation<sup>2</sup>. Both prokaryotic proteins, ELIC and GLIC form ion channels that are selective for cations over anions with poor discrimination among monovalent cations<sup>1,2</sup>, characteristics that resemble the conduction properties of the cation-selective branch

of the family that includes acetylcholine and serotonin receptors<sup>3,4</sup>. Here we present the X-ray structure of GLIC at 3.1 Å resolution. The structure reveals a conformation of the channel that is distinct from ELIC and that probably resembles the open state. In combination, both structures suggest a novel gating mechanism for pentameric ligand-gated ion channels where channel opening proceeds by a change in the tilt of the pore-forming helices.

GLIC was crystallized at pH 4, a condition where the open probability of the channel is high, and its structure was determined by molecular replacement followed by cyclic non-crystallographic symmetry averaging (Supplementary Figs 1–3, Supplementary Table 1). In its overall organization, the structure of GLIC closely resembles that of ELIC (Fig. 1): the extracellular domain encompasses 10 β-strands that are organized as a sandwich of two tightly interacting β-sheets, while the transmembrane domain folds into a bundle of four α-helices. As for other pLGICs of prokaryotic origin, an extended region between helices α3 and α4 is absent and replaced by a short loop<sup>5</sup>. The close correspondence between the ELIC and GLIC structures is emphasized in a superposition of their Cα positions (Fig. 1c). Despite the moderate sequence conservation (20% identical residues, Supplementary Fig. 4), both proteins are very similar on a structural level. Whereas the elements of secondary structure are highly conserved, spatial differences in the ligand-binding and the transmembrane domain indicate that the two



**Figure 1 | GLIC structure.** **a**, Ribbon representation of GLIC viewed from within the membrane with the extracellular solution above. The approximate membrane boundaries are indicated. **b**, Structure of the pentameric channel viewed from the extracellular side. **c**, Superposition of GLIC (green) and ELIC (orange, PDB 2vl0) subunits. The protein is shown as a Cα trace, the subunit is viewed from within the membrane. The superposition was generated by a least squares fit of the Cα positions of the respective α1 helices of the five subunits of the pentameric channel. Figures 1–4 were prepared with DINO (<http://www.dino3d.org>).

<sup>1</sup>Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland.



proteins adopt two distinct conformations of the ion conduction pore (Fig. 2, Supplementary Fig. 5).

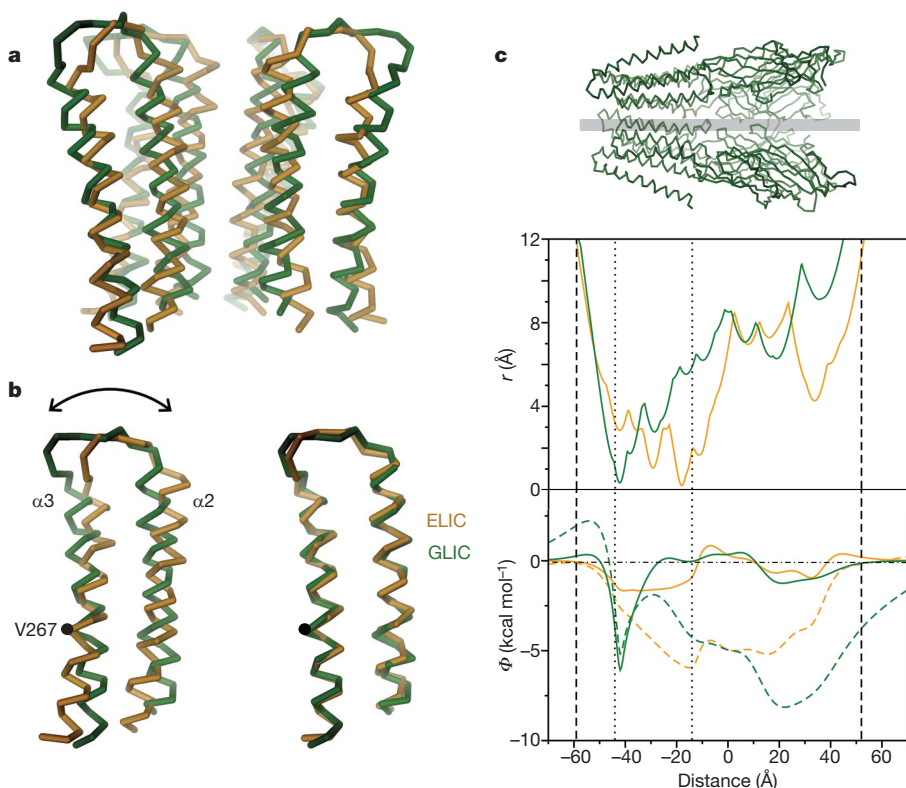
The conformational differences between the two ion channels are manifested in their pore geometry: whereas the transmembrane pore of ELIC is constricted on its extracellular side, the equivalent region of GLIC shows a funnel-shaped opening with a linearly decreasing diameter that places its narrowest part at the intracellular entry of the channel (Fig. 2c). At this constricted position of GLIC, the ELIC pore is dilated to a water-filled cavity of about 6 Å in diameter. The mismatch in the pore geometry results from differences in the orientation of the  $\alpha 2$  and  $\alpha 3$  helices. With respect to ELIC, both helices in GLIC have rotated as rigid unit by about  $9^\circ$  around an axis that intersects with residue Val 267 and that runs parallel to the membrane (Fig. 2a). This rotation results in an outward movement of the helix pair away from the pore axis on the extracellular side and an inward movement towards the axis at the intracellular entry of the channel. In contrast to the large differences in the position of the  $\alpha 2$  and  $\alpha 3$  helices, the location of helix  $\alpha 1$  that connects the transmembrane pore to the extracellular domain is virtually unchanged (Fig. 1c, Supplementary Fig. 5).

A comparison of the pore region of the nicotinic acetylcholine receptor (nAChR)—which was modelled into electron density at 4 Å resolution—with the corresponding region in the two bacterial channels reveals interesting similarities: the extracellular part of the nAChR pore resembles the funnel-shaped structure of GLIC, whereas the diameter in the intracellular half is closer to ELIC (Supplementary Fig. 6). Although the nAChR structure has been proposed to represent a non-conducting conformation<sup>6,7</sup>, as it was obtained in the absence of ligands, its relation to the two conformations of bacterial pentameric ligand-gated ion channels (pLGICs) at high resolution remains ambiguous.

Although the shape of the transmembrane channel in GLIC differs, the chemical features that are common to members of this branch of the family are preserved (Fig. 3)<sup>3</sup>. As in other cation-selective pLGICs, the extracellular half of the pore is dominated by hydrophobic residues, some of which carry bulky side chains. In contrast to the outer half, the intracellular part is lined by polar side chains of serine and

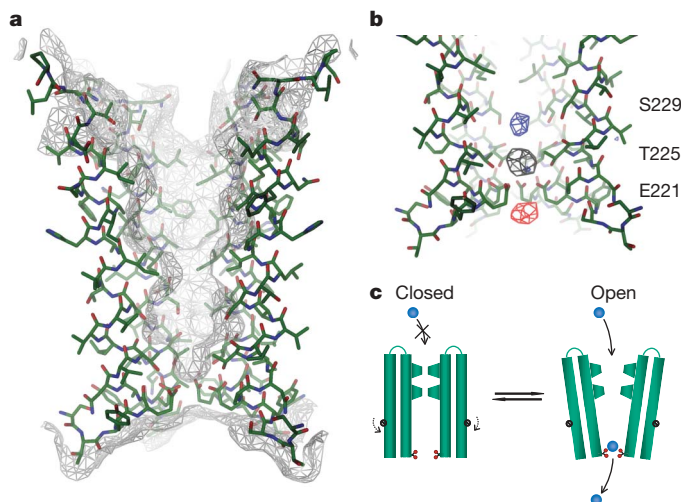
threonine residues that provide a hydrophilic character to this region. Both ends of the transmembrane channel are framed by rings of acidic residues that are located at the membrane boundary and that contribute to the overall negative electrostatic environment in the pore (Fig. 2c). Residual electron density in the water-filled channel indicates the presence of partially ordered solvent molecules (Supplementary Fig. 7). A ring of glutamate residues on the intracellular pore entry is remarkable. Equivalent residues in the nAChR, termed the ‘intermediate ring of charges’, have previously been identified as major determinants of conductance and cation selectivity<sup>8,9</sup>. In GLIC, the side chains of the five glutamate residues that are well defined in the electron density project into the lumen of the pore and form a ring of interacting ionizable groups (Fig. 3a, Supplementary Fig. 3b). They define the narrowest part of the channel and resemble a narrow selectivity filter, akin to other selective ion channels<sup>10</sup>, that might directly interact with ions which have stripped part of their hydration shell. When investigating crystals that were soaked in solutions containing the permeant cations  $\text{Rb}^+$ ,  $\text{Cs}^+$  and the divalent cation  $\text{Zn}^{2+}$ , we were able to locate electron density of bound ions in the region intra- and extracellular to the constriction, thus emphasizing its potential role for conductance and ion selectivity in GLIC (Fig. 3b, Supplementary Table 1).

The tight interactions of the five glutamate side chains defining the pore constriction might impede the free diffusion of ions. Because in a cellular context the intracellular part of the open channel faces the physiological pH of the cytoplasm, it is possible that the protonation state of these residues is shifted at low pH and that the interactions observed in the electron density are a consequence of the crystallization conditions. When investigated by excised inside-out patch-clamp recording, the current passed by the channel decreases upon lowering the intracellular pH, as expected when the net charge of the protein becomes more positive (Supplementary Fig. 1c). To investigate the extent to which the observed pore structure might be influenced by local interactions, we have determined the structure of the mutant E221A at 3.5 Å. The decreased cation selectivity of the mutant emphasizes the critical role of this residue for ion conduction (Supplementary Fig. 1d). The structure of the mutant, which



**Figure 2 | Comparison of the pore region.**

**a**, View of the C $\alpha$  trace of the  $\alpha 2$ – $\alpha 3$  helices. The respective regions of the superimposed structures of GLIC (green) and ELIC (orange) are shown. The front subunit is removed for clarity. The view is from within the membrane. **b**, Conformation of the  $\alpha 2$ – $\alpha 3$  helix pair of a single subunit as observed in the structures (left) and after a least squares fit of the C $\alpha$  positions of the respective helix pair of ELIC onto GLIC (right). The rotation axis is indicated. **c**, Pore radius ( $r$ ) and electrostatic potential ( $\phi$ ) in GLIC and ELIC. Orientation of GLIC is shown above; vertical dashed lines, molecular boundaries; vertical dotted lines, transmembrane region. Top plot,  $r$  of GLIC (green) and ELIC (orange). Bottom plot,  $\phi$  along the pore axis of GLIC (green) and ELIC (orange), as calculated from a numerical solution of the linearized Poisson–Boltzmann equation at 0 and 150 mM monovalent salt concentration (respectively dashed and solid lines).



**Figure 3 | Pore structure and ion binding.** **a**, View of the  $\alpha 2$  helices of GLIC defining the pore region. The front subunit is removed for clarity. The molecular surface is shown as grey mesh. **b**, Intracellular part of the pore region. Shown are the isomorphous difference electron density of  $\text{Cs}^+$  (grey, contoured at  $4.5\sigma$ ) and  $\text{Rb}^+$  (blue, contoured at  $3.5\sigma$ ), and the anomalous difference electron density of  $\text{Zn}^{2+}$  (red, contoured at  $4.5\sigma$ ). The position of selected residues is indicated. **c**, Schematic representation of the pore opening mechanism. The  $\alpha 2/\alpha 3$  helices of two subunits in the closed (left) and open (right) conformation are shown. The ion-coordinating glutamate residues are shown in red, the permeating ions in blue. The axis of rotation (black) in each subunit is indicated.

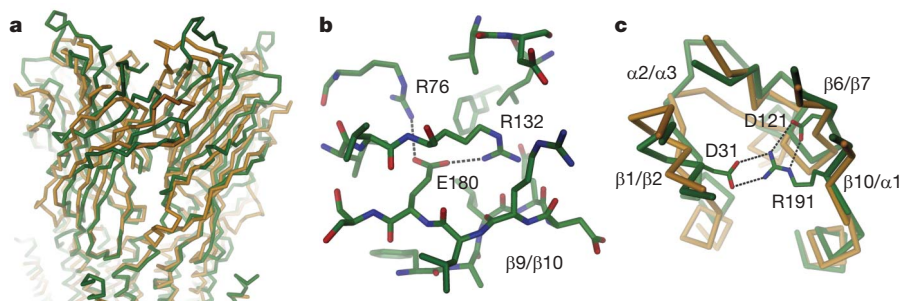
crystallized in similar conditions as wild type, showed an unaltered overall conformation of the protein backbone with a pore where the intracellular constriction is removed (Supplementary Fig. 8). A similar pore structure would be obtained in a case where the glutamate side chains change their conformation (Supplementary Fig. 8b, c). Our results thus suggest that the overall conformation of the pore-forming helices is not perturbed by local interactions of the partly protonated residues at low pH, and that minor changes in the side-chain conformation are sufficient to increase the pore radius. The conical aqueous pore structure and the ability to bind ions at the narrow intracellular constriction thus suggest that the structure of GLIC either depicts a conducting state of a pLGIC or a conformation that is very close to a conducting state (Fig. 3c).

Although GLIC shares the hallmarks of pLGICs, the ligand that triggers its channel opening is unusual. GLIC has been shown to be activated by a decrease in the extracellular pH (ref. 2). It is thus conceivable that the titration of a group in the consensus ligand-binding region causes conformational transitions in the extracellular domain that are similar to those induced in response to the binding of neurotransmitters in other members of the family<sup>11</sup>. The ligand-binding site of pLGICs is located at the interface between two subunits, and is covered by the strands  $\beta 9$  and  $\beta 10$  (Fig. 4). When compared to ELIC and acetylcholine-binding protein (AChBP) this

site shows distinctive features. The structure around  $\beta 9$  is only weakly conserved, particularly in the loop region connecting  $\beta 9$  and  $\beta 10$  (which is seven residues shorter than in ELIC) (Fig. 4a, Supplementary Fig. 4). Whereas in ELIC the loop is mobile and thus only poorly defined in the electron density, its conformation in GLIC is well defined with residues being involved in a variety of interactions, some of which involve ionizable residues (Fig. 4b, Supplementary Fig. 3c). Several of these interactions are unique to GLIC, such as a salt bridge between Arg 76 and Glu 180 at the position where the quaternary ammonium is bound in AChBP<sup>12,13</sup>. The tight interactions in the loop region of GLIC resemble the conformation that has been observed in AChBP in the presence of its substrate analogue carbamylcholine<sup>13</sup>, except that the pocket for the ligand found in AChBP and in the respective position in ELIC is filled with protein side chains (Supplementary Fig. 9). The structural features of the ligand-binding region are thus in accordance with an activated ligand-occupied state of the domain.

Conformational differences in the conserved regions of the ligand-binding domain hint towards transitions that lead to channel opening. When compared to ELIC, parts of the ligand-binding domain in GLIC have undergone considerable structural rearrangements. The movements are most pronounced for residues in one of the two  $\beta$ -sheets constituting the extracellular domain that was previously termed the 'inner sheet', part of which has tilted by about  $5^\circ$  towards the membrane plane (Fig. 4a). The structural differences between ELIC and GLIC also affect the loop regions connecting the strands  $\beta 1/\beta 2$  and  $\beta 6/\beta 7$  at the boundary between the cytoplasmic domain and the pore. Both regions contact the  $\alpha 2/\alpha 3$  loop in the transmembrane domain and have previously been identified to play a critical role in channel opening<sup>14,15</sup>. Yet, the extent of the movement in the extracellular domain is small compared to the large conformational difference in the adjacent region of the pore domain, as if the breaking of a critical interaction in the interface would have allowed the relaxation of a strained conformation of the channel (Fig. 1c, Supplementary Figs 3d, 10). Within the extracellular domain, the GLIC structure shows the relationship between conserved residues that are critical for channel opening. These residues include Arg 191 at the end of  $\beta 10$  that forms a pair of salt bridges with Asp 31 in the  $\beta 1/\beta 2$  turn and Asp 121 in the  $\beta 6/\beta 7$  loop (which is part of the Cys-loop signature sequence, Fig. 4c). Mutations that remove either of the two interactions in the nAChR and other pLGICs drastically decrease the open probability of the channel, thus emphasizing the importance of the respective loop regions for the transduction of conformational signals from the ligand-binding site to the pore domain<sup>14–17</sup>.

The pore structure of GLIC is in striking accordance with electrophysiological investigations on the conducting conformation of the nAChR. Different studies have previously predicted a funnel-shaped channel with a short constriction on the intracellular side that harbours a single ion-binding site at the narrowest part of the pore<sup>18,19</sup>. Whereas the structure of ELIC in a non-conducting conformation showed a pore that was constricted on the outside and lacked binding sites for ions in the pore, the structure of GLIC now displays the predicted features. The structure also allows a potential explanation



**Figure 4 | Extracellular domain.** **a**, View of the ligand-binding domains in the superimposed structures of GLIC (green) and ELIC (orange). Proteins are shown as  $\text{C}\alpha$  representation. **b**, Ligand-binding region of GLIC. Residues in the vicinity of the  $\beta 9/\beta 10$  loop are shown. The view is from the extracellular side. Selected residues and their interactions are indicated. **c**,  $\text{C}\alpha$  representation of the interface region between the extracellular domain and the pore in GLIC (green) and ELIC (orange) viewed from the extracellular side. Selected residues in GLIC are shown as sticks, ionic interactions are indicated by dotted lines.



for the dominant role of the residues constituting the 'intermediate ring of charges' in channel function<sup>8,9</sup>. These residues, which correspond to the ion-coordinating glutamates in GLIC, and that are in the same position in the homopentameric  $\alpha$ -7 nAChR and one helix turn towards the cytoplasm in the muscle nAChR, have been identified to exert a prevailing influence on ion conduction and cation selectivity<sup>20–22</sup>. The GLIC structure now suggests that the negatively charged side-chains directly interact with the partly desolvated permeant cations, rather than merely contributing to the overall negative electrostatics in the intracellular part of the pore. The analysis of the electrostatic potential of GLIC shows a pronounced energy minimum near the intracellular entry of the open channel that is decreased in the closed channel and that was previously observed in a functional study that investigated the binding of positive charged MTS (methanethiosulphonate) reagents to the nAChR (Fig. 2c)<sup>23,24</sup>.

The comparison of the two high resolution structures of pLGICs gives structural insight into the mechanisms underlying pore opening. These mechanisms have been ambiguous so far, and have mainly been derived from electron microscopy images at low resolution<sup>25</sup>. Although previous studies have proposed that channel opening proceeds by a rotation of the pore-lining helices around their helix axis<sup>6</sup>, the results of our study suggest a different mechanism. Assuming that the structures of ELIC and GLIC depict the closed and open conformations of a pLGIC pore, opening occurs by a change in the tilt of the helices  $\alpha$ 2 and  $\alpha$ 3 that move as a rigid body around an axis that runs parallel to the membrane and that intersects with a residue in  $\alpha$ 3 about two-thirds across the transmembrane channel (Fig. 3c). This mechanism is in accordance with previously recorded functional data<sup>26–28</sup>. Our study thus provides a first structural view at high resolution into how a pLGIC may open and selectively conduct ions.

## METHODS SUMMARY

GLIC was expressed in *Escherichia coli* with its amino terminus fused to maltose-binding protein (MPB) that was preceded by a signal sequence (PelB). The protein was purified from isolated membranes in the detergents n-dodecyl- $\beta$ -D-maltoside or n-tridecyl- $\beta$ -D-maltoside. MPB was cleaved during purification by proteolytic digestion at a specific site located between MBP and the channel protein. Crystals were grown at pH 4.0 with addition of 9% PEG 4000, 225 mM  $(\text{NH}_4)_2\text{SO}_4$  and 0.5 mg ml<sup>-1</sup> *E. coli* lipids. Data at 3.1 Å resolution were collected at the X06-SA beamline at the Swiss Light Source (SLS) of the Paul Scherrer Institute (PSI) on a Pilatus detector (Dectris) (Supplementary Table 1). The crystals were of space group C2 with one homo-pentameric channel in the asymmetric unit (Supplementary Fig. 2). The structure was determined by molecular replacement using as search model a modified structure of ELIC, with all side chains truncated and helix  $\alpha$ 2 and loop regions at the interface between the ligand-binding domain and the channel removed. Phases were improved and extended by cyclic fivefold NCS averaging. The final electron density was of very high quality and devoid of model bias (Supplementary Fig. 3). All subunits are structurally very similar and were refined maintaining strong constraints throughout (Supplementary Table 1). The final structure that encompasses residues 7–316 is well refined with good stereochemistry and with no outliers in disallowed regions of the Ramachandran plot.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 23 June; accepted 26 September 2008.

Published online 5 November 2008.

- Hilf, R. J. & Dutzler, R. X-ray structure of a prokaryotic pentameric ligand-gated ion channel. *Nature* **452**, 375–379 (2008).
- Bocquet, N. *et al.* A prokaryotic proton-gated ion channel from the nicotinic acetylcholine receptor family. *Nature* **445**, 116–119 (2007).
- Karlin, A. Emerging structure of the nicotinic acetylcholine receptors. *Nature Rev. Neurosci.* **3**, 102–114 (2002).
- Adams, D. J., Dwyer, T. M. & Hille, B. The permeability of endplate channels to monovalent and divalent metal cations. *J. Gen. Physiol.* **75**, 493–510 (1980).
- Tasneem, A., Iyer, L. M., Jakobsson, E. & Aravind, L. Identification of the prokaryotic ligand-gated ion channels and their implications for the mechanisms and origins of animal Cys-loop ion channels. *Genome Biol.* **6**, R4 (2004).

- Miyazawa, A., Fujiyoshi, Y. & Unwin, N. Structure and gating mechanism of the acetylcholine receptor pore. *Nature* **423**, 949–955 (2003).
- Unwin, N. Refined structure of the nicotinic acetylcholine receptor at 4 Å resolution. *J. Mol. Biol.* **346**, 967–989 (2005).
- Imoto, K. *et al.* Rings of negatively charged amino acids determine the acetylcholine receptor channel conductance. *Nature* **335**, 645–648 (1988).
- Konno, T. *et al.* Rings of anionic amino acids as structural determinants of ion selectivity in the acetylcholine receptor channel. *Proc. R. Soc. Lond. B* **244**, 69–79 (1991).
- Zhou, Y., Morais-Cabral, J. H., Kaufman, A. & MacKinnon, R. Chemistry of ion coordination and hydration revealed by a K<sup>+</sup> channel-Fab complex at 2.0 Å resolution. *Nature* **414**, 43–48 (2001).
- Sine, S. M. & Engel, A. G. Recent advances in Cys-loop receptor structure and function. *Nature* **440**, 448–455 (2006).
- Brejci, K. *et al.* Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature* **411**, 269–276 (2001).
- Celie, P. H. *et al.* Nicotine and carbamylcholine binding to nicotinic acetylcholine receptors as studied in AChBP crystal structures. *Neuron* **41**, 907–914 (2004).
- Lee, W. Y. & Sine, S. M. Principal pathway coupling agonist binding to channel gating in nicotinic receptors. *Nature* **438**, 243–247 (2005).
- Sala, F., Mulet, J., Sala, S., Gerber, S. & Criado, M. Charged amino acids of the N-terminal domain are involved in coupling binding and gating in  $\alpha$ 7 nicotinic receptors. *J. Biol. Chem.* **280**, 6642–6647 (2005).
- Schofield, C. M., Jenkins, A. & Harrison, N. L. A highly conserved aspartic acid residue in the signature disulfide loop of the  $\alpha$ 1 subunit is a determinant of gating in the glycine receptor. *J. Biol. Chem.* **278**, 34079–34083 (2003).
- Jha, A., Cadogan, D. J., Purohit, P. & Auerbach, A. Acetylcholine receptor gating at extracellular transmembrane domain interface: The cys-loop and M2–M3 linker. *J. Gen. Physiol.* **130**, 547–558 (2007).
- Dani, J. A. Open channel structure and ion binding sites of the nicotinic acetylcholine receptor channel. *J. Neurosci.* **9**, 884–892 (1989).
- Dani, J. A. & Eisenman, G. Monovalent and divalent cation permeation in acetylcholine receptor channels. Ion transport related to structure. *J. Gen. Physiol.* **89**, 959–983 (1987).
- Galzi, J. L. *et al.* Mutations in the channel domain of a neuronal nicotinic receptor convert ion selectivity from cationic to anionic. *Nature* **359**, 500–505 (1992).
- Corringer, P. J. *et al.* Mutational analysis of the charge selectivity filter of the  $\alpha$ 7 nicotinic acetylcholine receptor. *Neuron* **22**, 831–843 (1999).
- Gunthorpe, M. J. & Lummis, S. C. Conversion of the ion selectivity of the 5-HT(3a) receptor from cationic to anionic reveals a conserved feature of the ligand-gated ion channel superfamily. *J. Biol. Chem.* **276**, 10977–10983 (2001).
- Pascual, J. M. & Karlin, A. State-dependent accessibility and electrostatic potential in the channel of the acetylcholine receptor. Inferences from rates of reaction of thiosulfonates with substituted cysteines in the M2 segment of the  $\alpha$ 1 subunit. *J. Gen. Physiol.* **111**, 717–739 (1998).
- Wilson, G. G., Pascual, J. M., Brooijmans, N., Murray, D. & Karlin, A. The intrinsic electrostatic potential and the intermediate ring of charge in the acetylcholine receptor channel. *J. Gen. Physiol.* **115**, 93–106 (2000).
- Unwin, N. Acetylcholine receptor channel imaged in the open state. *Nature* **373**, 37–43 (1995).
- Paas, Y. *et al.* Pore conformations and gating mechanism of a Cys-loop receptor. *Proc. Natl Acad. Sci. USA* **102**, 15877–15882 (2005).
- Cymes, G. D., Ni, Y. & Grosman, C. Probing ion-channel pores one proton at a time. *Nature* **438**, 975–980 (2005).
- Cymes, G. D. & Grosman, C. Pore-opening mechanism of the nicotinic acetylcholine receptor evinced by proton transfer. *Nature Struct. Mol. Biol.* **15**, 389–396 (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank B. Blattmann and C. Stutz-Ducommun for assistance with crystal screening, C. Schulze-Briese and the staff of the X06SA beamline for support during data collection, the protein analysis group at the functional genomics centre of the University of Zurich for help with mass spectrometry, R. MacKinnon for comments on the manuscript and members of the Dutzler laboratory for help in all stages of the project. Data collection was performed at the Swiss Light Source of the Paul Scherrer Institute. The research leading to these results was supported by a grant from the National Center for Competence in Research (NCCR) in Structural Biology and by an EC FP7 grant for the EDICT consortium (HEALTH-201924). R.J.C.H. is affiliated with the Molecular Life Sciences Ph.D. programme of the University/ETH Zurich.

**Author Contributions** R.D. and R.J.C.H. designed the project. R.J.C.H. carried out all experiments. R.D. assisted in data collection and structure determination. R.D. and R.J.C.H. jointly wrote the manuscript.

**Author Information** Coordinates of GLIC have been deposited in the Protein Data Bank under accession number 3EHZ (WT), and 3EIO (E221A). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to R.D. ([dutzler@bioc.uzh.ch](mailto:dutzler@bioc.uzh.ch)).

## METHODS

**Expression and purification.** The gene encoding GLIC was cloned into a modified pET26b vector (Novagen) where it was fused to the carboxy terminus of MBP, generating a construct with the pelB signal sequence followed by a (His)<sub>10</sub>-tag, MBP, a *Herpes simplex* (HRV) 3C protease site (GE Healthcare) and the GLIC sequence. BL21 (DE3) cells transformed with this construct were grown at 37 °C in TB medium containing 50 mg l<sup>-1</sup> kanamycin to an A<sub>600</sub> of ~1.6–1.8. Expression was induced by addition of 0.2 mM IPTG overnight at 20 °C. All following steps were carried out at 4 °C. Cells were harvested and lysed with an Emulsiflex high pressure homogenizer (Avestin) in 50 mM potassium phosphate (pH 8.0), 150 mM NaCl (buffer A), with the addition of 1 mM phenylmethyl sulphonyl fluoride (PMSF), 20 µg ml<sup>-1</sup> DNase I, 1 mg ml<sup>-1</sup> lysozyme, 1 µg ml<sup>-1</sup> pepstatin and 1 µg ml<sup>-1</sup> leupeptin. The lysate was cleared by low-spin centrifugation. Membranes were isolated by ultracentrifugation and the proteins were extracted in buffer A containing 2% n-dodecyl-β-D-maltoside (DDM, Anatrace). After centrifugation, the protein was purified by affinity chromatography on a Ni-NTA column (Qiagen). The purified MBP-GLIC-fusion protein was digested with HRV 3C protease to cleave the His<sub>10</sub>-MBP protein and dialysed against 10 mM potassium phosphate (pH 8.0), 150 mM NaCl and either 0.5 mM DDM or 0.2 mM n-tridecyl-β-D-maltoside (TDM, buffer B). His<sub>10</sub>-MBP was subsequently removed from solution by binding to Ni-NTA resin. GLIC was concentrated and subjected to gel filtration on a Superdex 200 column (GE Healthcare) in buffer B. The protein peak corresponding to the GLIC pentamer was pooled and concentrated to 10 mg ml<sup>-1</sup>, and used for crystallization.

**Crystallization.** GLIC was crystallized in sitting drops at 4 °C. Crystals were obtained by mixing protein containing additional 0.5 mg ml<sup>-1</sup> *E. coli* polar lipids (Avanti Polar Lipids) in a 1:1 ratio with reservoir solution containing 225 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 50 mM sodium acetate (pH 4.0), 9–12% PEG 4000. For the crystallization of the GLIC mutant E221A, the reservoir solution contained 500 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 50 mM sodium acetate (pH 4.0), 9–12% PEG 4000. For cryoprotection, the crystals were transferred into mother liquor containing 30% ethyleneglycol and flash-frozen in liquid propane. For soaking in 100 mM of Rb<sup>+</sup> and Zn<sup>2+</sup>, crystals were incubated in mother liquor where NaCl was replaced by the respective chloride salt; for soaking in high (250 mM) concentration of Cs<sup>+</sup> ions, crystals were incubated in mother liquor where (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> and NaCl were replaced by Cs<sub>2</sub>SO<sub>4</sub>.

**Structure determination.** Data sets were collected on frozen crystals on the X06SA beamline at the Swiss Light Source (SLS) of the Paul Scherrer Institut (PSI) on a PILATUS detector (Dectris). The data were indexed, integrated and scaled with the programs XDS<sup>29</sup>, MOSFLM<sup>30</sup> and SCALA<sup>31</sup> and further processed with CCP4 programs<sup>32</sup> (Supplementary Tables 1 and 2). Initial phases were obtained by molecular replacement with PHASER<sup>33</sup> using a partial structure of ELIC as search model. All side chains of the ELIC pentamer were truncated to alanine and helix α-2 and the loop regions in the interface between the two domains were removed. Phases were improved and extended to 3.1 Å by fivefold NCS symmetry averaging in DM<sup>34</sup>. The model was built in O<sup>35</sup> and initially refined maintaining strict fivefold NCS constraints in CNS<sup>36</sup> and PHENIX<sup>37</sup>. In later stages, the strict constraints were loosened and restraint individual B-factors were refined. R and R<sub>free</sub> were monitored throughout. R<sub>free</sub> was calculated by selecting 5% of the reflection data in thin slices that were omitted in refinement. The final model has R/R<sub>free</sub> values of 23.8% and 26.6%, good geometry and no outliers in the Ramachandran plot (Supplementary Table 1). The pore radii were calculated with the program HOLE<sup>38</sup>. The molecular surface was calculated with MSMS<sup>39</sup>.

**Poisson–Boltzmann calculations.** Electrostatic potential was calculated by solving the linearized Poisson–Boltzmann equation in CHARMM<sup>40,41</sup> on a 110 × 110 × 150 Å grid (1 Å grid spacing) followed by focusing on a 90 Å × 90 Å × 140 Å grid (0.5 Å grid spacing). Partial protein charges were derived from the Param19 extended hydrogen atom force field. Polar hydrogen

positions were generated in CHARMM. The protein was assigned a dielectric constant of 2. Its transmembrane region was embedded in a 30-Å-thick slab (ε = 2) representing the membrane that contained a cylindrical hole around the water-filled pore region and was surrounded by an aqueous environment (ε = 80). Calculations were repeated with mobile ion concentrations of 0 and 150 mM of monovalent salt in the solvent region excluding the transmembrane channel.

**Electrophysiology.** Constructs containing the gene of GLIC and the mutant E221A preceded by the signal sequence of the human α7-nAChR were cloned into the pTLN vector for expression in *Xenopus laevis* oocytes<sup>42</sup>. After linearization of the plasmid DNA by MluI, capped complementary RNA was transcribed with the Message Machine kit (Ambion) and purified with the RNeasy kit (Qiagen). For expression, 1–50 ng of RNA was injected into defolliculated oocytes. Two-electrode voltage-clamp measurements were performed 2–3 days after injection at 20 °C (OC-725B, Warner Instrument Corp.; Supplementary Fig. 1). Currents were recorded in bath solutions containing 130 mM NaCl, 2 mM KCl, 1.8 mM CaCl<sub>2</sub> and 5 mM MgCl<sub>2</sub>. The respective pH values were stabilized by addition of 10 mM Hepes (pH 7.0) or 10 mM of sodium citrate (pH 4.0–5.5). For inhibition studies 5 mM of TBA was added. Low salt solutions contained 30 mM NaCl, 100 mM mannitol, 2 mM KCl, 1.8 mM CaCl<sub>2</sub>, 1 mM MgCl<sub>2</sub> and 10 mM sodium citrate at pH 4.0. Inside-out patch-clamp measurements were recorded 3 days after injection with an Axopatch 200B amplifier (Axon Instruments; Supplementary Fig. 1c). Electrodes had a resistance of 1–3 MΩ. Bath solution contained 150 mM NaCl, 1 mM CaCl<sub>2</sub> and 5 mM MgCl<sub>2</sub>. The respective pH values were stabilized by addition of 10 mM Hepes (pH 7.0) or 10 mM of sodium citrate (pH 4.0, pH 5.0). The pipette solutions contained 150 mM NaCl, 1 mM EGTA and 5 mM MgCl and 10 mM sodium citrate pH 4.0.

29. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800 (1993).
30. Leslie, A. G. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* (No. 26, Daresbury Laboratory, 1992).
31. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
32. CCP4. Collaborative Computational Project Nr. 4. The CCP4 Suite: Programs for X-ray crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
33. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
34. Cowtan, K. An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* 34–38 (No. 31, Daresbury Laboratory, 1994).
35. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
36. Brunger, A. T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
37. Adams, P. D. et al. PHENIX: Building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
38. Smart, O. S., Neduvellil, J. G., Wang, X., Wallace, B. A. & Sansom, M. S. HOLE: A program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* **14**, 354–360, 376 (1996).
39. Sanner, M. F., Olson, A. J. & Spehner, J. C. Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers* **38**, 305–320 (1996).
40. Brooks, B. R. et al. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1983).
41. Im, W., Beglov, D. & Roux, B. Continuum solvation model: Electrostatic forces from numerical solutions to the Poisson–Boltzmann equation. *Comput. Phys. Commun.* **111**, 59–75 (1998).
42. Lorenz, C., Pusch, M. & Jentsch, T. J. Heteromultimeric CLC chloride channels with novel properties. *Proc. Natl Acad. Sci. USA* **93**, 13362–13366 (1996).



# naturejobs

**THE CAREERS  
MAGAZINE FOR  
SCIENTISTS**

**A**t a time when financial markets are shaky, science career opportunities uncertain and the plight of the serial postdoc as much of a concern as ever, this year's Postdoc Journal keepers offer hope. In the past few months two of them have earned permanent positions. One has embarked on the challenge of motherhood. The fourth has found a research position that, although immensely challenging, always enthalls her. Each sums up their experiences and future ambitions in online essays (see <http://tinyurl.com/57hgy7>).

Jon Yearsley, a veritable postdoc-aholic of ten years' experience, managed to live up to one of his New Year's resolutions from last year: to have a permanent position or no position at all. He recently landed a lecturing job in Dublin and was joined by his partner after much time apart. Of course, challenges remain: will his brand of interdisciplinary research attract funding? What about students?

New responsibilities have proven daunting for Zachary Lippman, who has just moved to Cold Spring Harbor Laboratory in New York as an assistant professor after a postdoc in Israel. He knows he must generate preliminary data and write grant proposals, but he remains wistful about his past experiences and colleagues.

Aliza le Roux started the year with a research trek of epic proportions. She went from her native South Africa to her new principal investigator's lab at Michigan State University before, in short order, heading to Ethiopia for her new postdoc. She is fascinated every day by the natural soap opera performed by a troop of baboons she studies there. Le Roux loves her job, she says unabashedly. She gets to play with animals and chase her own intellectual questions. But she worries that her work might make a 'normal' life a difficult prospect.

Amanda Goh, on the other hand, has taken a step towards that so-called normal life. Goh became quite frustrated with her lab work this year; she contemplated starting a completely new project. Now, as she prepares for the birth of her first child, she already recognizes that the careful planning she strives for in the lab may be even more difficult to carry out in the nursery.

**Gene Russo is editor of *Naturejobs*.**

## CONTACTS

**Editor:** Gene Russo

**Assistant editor:** Karen Kaplan

e-mail: [naturejobseditor@nature.com](mailto:naturejobseditor@nature.com)

## European Head Office, London

The Macmillan Building,

4 Crinan Street, London N1 9XW, UK

Tel: +44 (0) 20 7843 4961

Fax: +44 (0) 20 7843 4996

e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

## European Sales Manager:

Andy Douglas (4975)

e-mail: [a.douglas@nature.com](mailto:a.douglas@nature.com)

## Assistant European Manager:

Nils Moeller (4953)

## Natureevents:

Ghizlaine Ababou (+44 (0) 20 7014 4015)

e-mail: [g.ababou@nature.com](mailto:g.ababou@nature.com)

## Southwest UK/RoW:

Alexander Ranken (4944)

## Northeast UK/Ireland:

Matthew Ward (+44 (0) 20 7014 4059)

## France/Switzerland/Belgium:

Muriel Lestringuez (4994)

## Scandinavia/Spain/Portugal/Italy:

Evelina Rubio-Hakansson (4973)

## North Germany/The Netherlands/Eastern

Europe: Kerstin Vincze (4970)

## South Germany/Austria:

Hildi Rowland (+44 (0) 20 7014 4084)

## Advertising Production Manager:

Stephen Russell

To send materials use London address above.

Tel: +44 (0) 20 7843 4816

Fax: +44 (0) 20 7843 4996

e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

**Naturejobs web development:** Tom Hancock

**Naturejobs online production:** Dennis Chu

## US Head Office, New York

75 Varick Street, 9th Floor,

New York, NY 10013-1917

Tel: +1 800 989 7718

Fax: +1 800 989 7103

e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

**US Sales Manager:** Ken Finnegan

## India

Vikas Chawla (+91 1242881057)

e-mail: [v.chawla@nature.com](mailto:v.chawla@nature.com)

## Japan Head Office, Tokyo

Chiyoda Building, 2-37 Ichigayatamachi,

Shinjuku-ku, Tokyo 162-0843

Tel: +81 3 3267 8751

Fax: +81 3 3267 8746

## Asia-Pacific Sales Manager:

Ayako Watanabe (+81 3 3267 8765)

e-mail: [a.watanabe@natureasia.com](mailto:a.watanabe@natureasia.com)

## Business Development Manager, Greater

China/Singapore:

Gloria To (+852 2811 7191)

e-mail: [g.to@natureasia.com](mailto:g.to@natureasia.com)

# Gravitational astronomy 101

It's a bit of a shock.

**David Blair**

I want to start this course with a brief review of the history of gravitational astronomy, starting from the Eta Carinae supernova of 2009. You will be in mid-career when the shock front hits, and a good understanding of GA will be useful for anyone remotely involved in the Shield Project.

The UN declared 2009 to be an International Year of Astronomy, in honour of the 400th anniversary of Galileo's first observations with a telescope. Just 5 years before Galileo discovered craters on the Moon and Jupiter's moons, Kepler had recorded a supernova. From that date until 2009, for the entire time since the invention of the telescope, there had been a complete absence of supernovae in our Galaxy. Astronomers thought this very unfair. If you had been born in the year 1000 you could have seen two supernovae in your lifetime, and the same if you had been born in 1550. But the unfairness of the 400-year absence is nothing compared with the unfairness of the Eta Carinae supernova. For Earth, the coming shock is a one-in-a-million-year event.

But I digress. Let's focus on the gravitational astronomy. Over the years before the EC supernova, there had been many years of painstaking development of gravitational-wave detectors. The world detector array that now runs autonomously was in its infancy. There was not even one Southern Hemisphere detector. The angular resolution (which we will derive in the next lecture) was lousy.

The EC supernova was a major embarrassment to science. Even though electromagnetic astronomers had been predicting it for half a century, no one noticed when it actually happened. The gravity signal was so different from any predicted waveform that it too was only identified months after the event. So the most momentous astronomical event in human history — one that, as you all know, threatens to put an end to human history — was ignored for days, until a few amateurs noticed the changes in Eta Carinae. The trouble was that Eta Carinae was very special and very rare: one star among 100 billion that was more than 100 times as heavy as our Sun. Its light was blanketed by the nebula and the prompt  $\gamma$ -ray burst lit up only a corner of the nebula.

Getting an accurate distance to EC was the key to predicting the shock arrival, which as you all know is expected in March '81. Like all gravity signals, the EC

signal gave us its distance, which is 7,239 light years. You would think that was sufficiently far away for Earth to be safe. To see why it's not, you have to understand geons and strong-field nonlinear coupling in supersymmetric loop quantum gravity.

Geons were discovered theoretically by John Archibald Wheeler in 1955. They are gravitationally bound entities of pure gravitational energy. Most astrophysicists believe that they could not actually exist, despite Gell-Mann's totalitarian principle, which says *if it is not prohibited then it is compulsory*. Geons are not prohibited by the laws of physics so they must be out there. But Wheeler did not consider gravitino-neutralino coupling. This causes strong-field gravitational wavefronts to gravitationally condense into geon-like bubbles that propagate at about  $0.99c$ . By the end of the course you should be able to derive the threshold gravitational-wave amplitude for the creation of gravitino-neutralino geons.

The year 2009 broke the supernova drought but it also gave us the great flood that is on its way. And for physicists it broke the drought in our understanding of the Universe. In 2009, physics had been derailed by an impossibly complex theory they called string theory. It was like Ptolemy's theory of the Solar System that derailed astronomy for more than a millennium. While physicists used string theory to try to solve an apparent incompatibility between quantum mechanics and gravity, astronomers were invoking stuff they called 'dark matter' and 'dark energy' to explain galaxy rotation and cosmic expansion. Physics was in a mess. The key to the great synthesis was in the 2009 gravity-wave data. The breakdown of general relativity at the black-hole event horizon seems obvious today, but it took Otulu 20 years to figure it out.

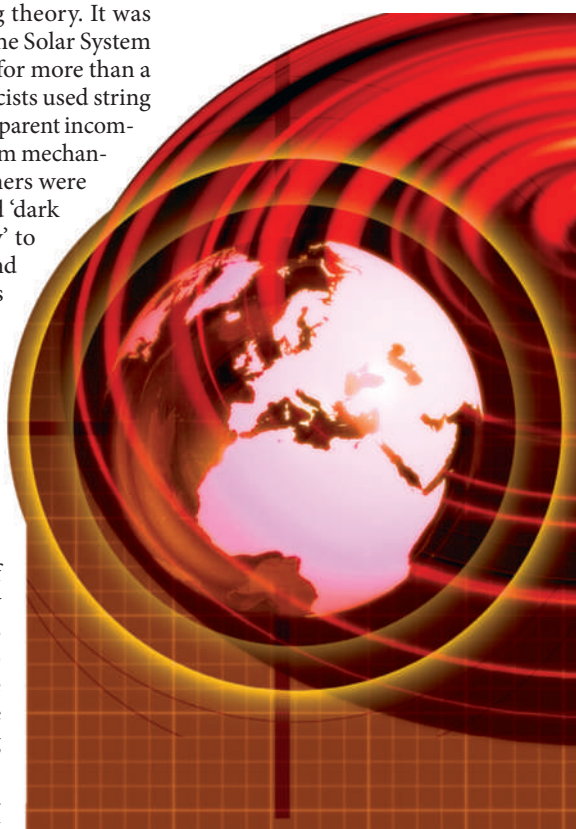
It is the nonlinearity of classical general relativity that causes the instabilities that turn shock fronts into cosmic bullets. This is the fundamental cause of the catastrophe of the coming shock and you must understand it in detail. At the end of the course I will expect

you to be able to derive both the gravitino-neutralino shock-front structure as well as its peak intensity and arrival time from the gravity-wave data, using Otulu's strong-coupling theory. You will need to understand how the scalar field drives the coherent decay of the shock when it interacts with stellar plasma. You will need to be familiar with the extragalactic data that confirmed the theory, and be able to show how propagating gravitino-neutralino-geon condensates explain both cosmic-expansion data and galaxy rotation. This will test your understanding of the great synthesis.

The absence of other galactic civilizations is probably due to their failure to understand gravitino-neutralino geons. I hope that some of you will go on to work for the Shield Project and make sure we do not follow in their footsteps.

Now, let's get to work. The first slide shows the Otulu equation ...

**David Blair is director of the Australian International Gravitational Research Centre at the University of Western Australia, a co-founder of the Gravity Discovery Centre ([www.gdc.asn.au](http://www.gdc.asn.au)) and Western Australia's Scientist of the Year. This story was inspired by discussions at CECS Chile.**



JACEY